ARL-TR-6922 • SEP 2015

US Army Research Laboratory

Pilot Variability Study for Federal Aviation Administration Health and Usage Monitoring Mock Certification

by Natasha C Bradley

NOTICES

Disclaimers

ARL-TR-6922 ● SEP 2015

US Army Research Laboratory

Pilot Variability Study for Federal Aviation Administration Health and Usage Monitoring Mock Certification

by Natasha C Bradley
Vehicle Technology Directorate, ARL

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

1. REPORT DATE *(DD-MM-YYYY)* September 2015	**2. REPORT TYPE** Final	**3. DATES COVERED (From - To)** October 2009–April 2014
4. TITLE AND SUBTITLE Pilot Variability Study for Federal Aviation Administration Health and Usage Monitoring Mock Certification		**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Natasha C Bradley		**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory ATTN: RDRL-VTM Aberdeen Proving Ground, MD 21005-5066		**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-6922
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The US Army Research Laboratory (ARL) conducted a Design of Experiment study to assess the variability in health and usage monitoring systems and to reduce the effect of any bias that may result due to changes in pilots. The objective of the statistical approach is to support Sikorsky Aircraft in evaluating their flight regime recognition (FRR) algorithms with respect to pilot variability. Damage data collected by the FRR algorithms are being used to accurately determine usage credits. Variability is defined as the dispersion of the data, which could be caused by the pilot’s individual techniques. In this analysis, ARL has provided an additional approach for Health and Usage Monitoring Systems certification as a comparison to Aeronautical Design Standard Handbook for Condition Based Maintenance Systems for US Army Aircraft (ADS-79D) recommendations. The statistical analysis shows that the null hypothesis (pilot variability) did not influence regime recognition algorithms. There is a 95% confidence level that pilot variability is not an issue.

15. SUBJECT TERMS

pilot variability, FAA mock certification, FAA HUMS program, pilots effect on HUMS, FAA usage credits

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Natasha C Bradley
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified	UU	20	410-278-7756

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.1

Contents

List of Figures

List of Tables

1. Introduction

The US Army Research Laboratory (ARL) conducted a Design of Experiment (DOE) study to assess the variability in health and usage monitoring systems (HUMS) and to reduce the effect of any bias that may result due to changes in pilots. The objective of the DOE is to support Sikorsky Aircraft in evaluating the flight regime recognition (FRR) algorithms with respect to pilot variability. Structural damage data collected by the HUMS and processed by the FRR algorithms are being used to accurately determine usage credits (a method of identifying the amount of usage an aircraft has accumulated over time, instead of the more traditional method of flight hours). Variability is defined as the dispersion in the data, which would be caused by the pilot's individual techniques. According to the Federal Aviation Administration, pilot variability must be addressed in order to certify a HUMS for use in aircraft. In this analysis, ARL has provided an additional approach for HUMS certification as a comparison to Aeronautical Design Standard Handbook for Condition Based Maintenance Systems for US Army Aircraft (ADS-79D) recommendations.[1]

The challenges faced in accessing pilot variability were flight time, pilot control of usage dependent factors, and number of pilots available. ARL researchers identified the top 3 such (critical) factors that affected (or were harmful) to the UH-60 helicopter structure. During flight maneuvers, these critical factors typically cause high loading and high stresses on the aircraft structure. For this study, the factors were 1) bank angle, 2) airspeed, and 3) rotorcraft weight. Bank angle and airspeed were easy for the pilots to control in flight; rotorcraft weight is set at takeoff. Critical regimes (in concert with critical factors) contain maneuvers that cause high loading on the rotorcraft structure during flight.

2. Design of Experiment Process

ADS-79D specifies how to evaluate pilot variability with regards to HUMS. The issue with evaluating pilot variability using ADS-79D is that method involves an excessive number of pilots and flight hours.[1] ARL's focus was to understand which regimes (or sequence of flight maneuvers) in the recognition process were critical. These critical regimes were then selected and flown by different pilots to evaluate the effect of pilot variability on FRR determined usage credits. ARL evaluated the components that made up a DOE. The 3 main components of a DOE are treatments (factors), experimental units, and responses. There were 4 treatments: bank angle; airspeed; rotorcraft weight; and most important for this study, different pilot. The nonpilot treatments were chosen because it is known that they have a significant effect on vibratory loads. The pilot effect is less well known. The experimental units were the different flight maneuvers that the pilots would fly. The responses were the vibratory load data that were collected. Analysis of each critical maneuver lead ARL to choose the turn maneuver as the most detrimental (critical) maneuver recognized by the regime recognition method being tested. The method chosen to review the data for significance after testing was the Analysis of Variance (ANOVA).

Before ANOVA can be employed, a factorial method could be applied to evaluate the structural loading data collected from specific flight maneuvers.[2] Using factorial analysis, it was determined that 16 maneuvers would be needed to test 4 factors twice at both high and low level.[2] This can be done by 2 pilots within only 4 flights. However, the ADS-79D Handbook recommends for each maneuver in each critical regime 3 pilots are needed, each performing 3 flights; this would total 9 flights.[1] Thus the application of the factorial analysis reduced the flight time and number of pilots by five-ninths and one-third, respectively.

As discussed earlier, the accompanying Figure displays how ANOVA fits into the process used to evaluate the pilot variability. To reiterate, the process involved first defining the critical factors (the factors tested were determined in collaboration with the US Army Communications-Electronics Research, Development and Engineering Center [CERDEC] pilots) and critical regimes and then formulating flight cards that contain the critical regimes, followed by flight testing, and finally ANOVA analysis of FRR data. In a standard factorial, each factor has at least 2 different levels to be tested. In particular, a high and low discrete (possible) value or "level" for each factor defined by Sikorsky. These levels are to represent the maximum and the minimum of the standard operation envelope. They were selected for each factor within the normal operating conditions.3 Levels in a factorial are important to help understand the possible combination of conditions across each factor. Section 3 will elaborate on how the flight scripts were developed.

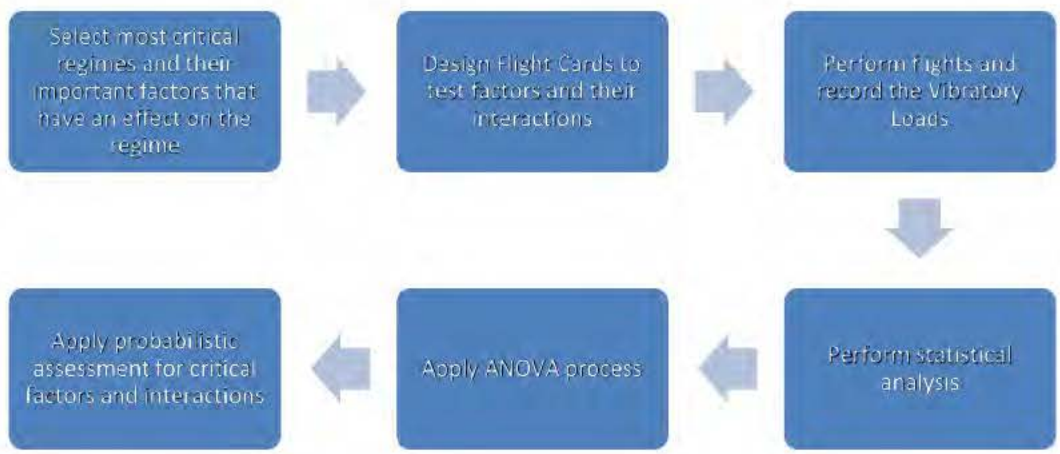

Figure DOE process using ANOVA for pilot variability study

3. Flight Scripts

The flight scripts were developed using the suggestions of the pilots, the rotorcraft limits, the maneuver order and the available flight time. Table 1 shows the selected flight maneuvers within a section of the entire flight script. The first flight was a clean configuration, meaning rotorcraft weight only, and the second flight included additional weight from stores. A flight card was made up for each maneuver. It consisted of a detailed description of the run number, the speed, and the conditions that would be met. Each of the flight maneuvers, 4 to the left and 4 to the right, were run in succession by the pilot and then copilot then repeated using a different configuration. Repeating the maneuvers in different directions maximized the amount of data collected because despite of a difference in direction, the maneuver was the same

Table 1   Sample flight test card where the turns are varied

Run	Description	Air Speed (±5kts)	Target Condition	Operator	Approx Duration (s)
44	Left-level turn, Entry/Steady/Recovery	80	35 °	Pilot	15
45	Left-level turn, Entry/Steady/Recovery	80	45 °	Pilot	15
46	Right-level turn, Entry/Steady/Recovery	80	35 °	Pilot	15
47	Right-level turn, Entry/Steady/Recovery	80	45 °	Pilot	15
48	Left-level turn, Entry/Steady/Recovery	100	35 °	Pilot	15
49	Left-level turn, Entry/Steady/Recovery	100	45 °	Pilot	15
50	Right-level turn, Entry/Steady/Recovery	100	35 °	Pilot	15
51	Right-level turn, Entry/Steady/Recovery	100	45 °	Pilot	15
52	Left-level turn, Entry/Steady/Recovery	80	35 °	Copilot	15
53	Left-level turn, Entry/Steady/Recovery	80	45 °	Copilot	15
54	Right-level turn, Entry/Steady/Recovery	80	35 °	Copilot	15
55	Right-level turn, Entry/Steady/Recovery	80	45 °	Copilot	15
56	Left-level turn, Entry/Steady/Recovery	100	35 °	Copilot	15
57	Left-level turn, Entry/Steady/Recovery	100	45 °	Copilot	15
58	Right-level turn, Entry/Steady/Recovery	100	35 °	Copilot	15
59	Right-level turn, Entry/Steady/Recovery	100	45 °	Copilot	15
60	Steady climb	120	1,000 fpm		10

4. ANOVA Analysis

ANOVA provided analysis on whether the results of the testing are statistically significant. The result would be called statistically significant only if it occurred by chance.[2] The ANOVA results were utilized to accept or reject the null hypothesis that the pilot-flying technique will not affect the damage accumulation recorded by the HUMS.

In particular, 4 different ANOVA analyses were conducted; the comparisons, done between treatments, were rotorcraft weight, airspeed, bank angle, and the variability of the pilots, and the variability of the pilots. The process included using sample variances to solve for the value of F. F is the comparison of variances found between the factors tested, to within the factors tested. F was compared to F-critical, which was the number that the test statistic must exceed to reject the null hypothesis and can be determined by F-distribution and tables.

$$F = \frac{MS_{Source\ of\ variation}}{MS_{Within}} . \tag{1}$$

The numerator is the Mean Squares (MS) for the variance between the treatments and the denominator is the MS for variance within the treatments.

$$MS = \frac{SS}{df} . \tag{2}$$

Where SS the sum of square and df is the degree of freedom (which is one less than the number of samples used when calculating the corresponding SS). The *SS* is calculated by

$$SS = \sum_{i=1}^{n}(x_i - \bar{x})^2, \tag{3}$$

where *n* is the number of samples, x_i are the individual sample values, and $\bar{x}$ is the mean of the sample.

The F-value is determined by the F-ratio, which is where *I* = number of factors compared and *J* = number of total samples. The p-value is the probability calculated from the ANOVA process; it was compared with the alpha (which is the probability of rejecting the null hypothesis given it is true) of the test. If the p-value is larger than 0.05 (alpha), then it would be safe (at the 95% confidence level) to accept the null hypothesis that states that pilot variability has no impact on the effectiveness of the regime recognition algorithms. For any p-value larger than 0.05, the null hypothesis must be rejected. Rejecting the null hypothesis does not mean that pilot variability has an impact, it just means that one cannot say with confidence that the pilot variability would not have an effect. In the case that the null hypothesis is rejected, subsequent testing will need to be performed.

5. Results

None of the test analyses showed any significant pilot variability with respect to the 95th percentile. This was enough information to accept the null hypothesis and as such, pilot variability has no effect on the accuracy of the regime recognition algorithms. As the pilot flew through each of the flight cards, a structural strain measurement was taken by the HUMS instrumentation and calibrated into load. These load measurements were the data collected and then used for the ANOVA. Table 2 shows the loads for each pilot as they flew through the different flight maneuvers on the flight card. These data were then summarized by finding the sum, average, and variance of each group, which is included in Table 3. Table 4 shows the ANOVA output as it is printed from Excel. When an ANOVA is completed on pilot flight data, the probability is greater than 0.05; therefore, the null hypothesis can be accepted. Table 5 shows the loads for each pilot with respect to the 2 different rotorcraft weight configurations. These data were then summarized by finding the sum, average, and variance of each pilot with respect to the rotorcraft configuration, which as shown in Table 6. In Table 7 the p-value for the sample is smaller than 0.05, which is a rejection of the null hypothesis. That variability was between the clean configuration and the extra weight configurations called External Stores Support System (ESSS), but this would not be considered pilot variability.

The variability in the sampling was expected because there was a large difference between having an empty rotorcraft and a rotorcraft that is weighed down with additional stores. This reasoning and the results of the integration between rotorcraft weight and pilot variability was enough evidence to accept the null hypothesis despite this aberration.

Table 2 Loading conditions from the pilot throughout flight

Loading Conditions (lbs)	
Pilot A	**Pilot B**
1,221	1,293
1,293	1,257
1,185	1,221
1,221	1,221
1,293	1,257
1,400	1,329
1,329	1,257
1,329	1,293
1,479	1,516
1,641	1,623
1,572	1,406
1,465	1,703
1,621	1,292
1,720	1,340
1,468	1,625
1,676	1,446

Table 3 ANOVA single factor summary for pilot

Groups	Count	Sum (lbs)	Average (lbs)	Variance
Pilot A	16	22,913	1,432.06	30,467.396
Pilot B	16	22,079	1,379.94	24,664.196

Table 4 ANOVA single factor output for pilot

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	21,736.13	1	21,736.13	0.7885	0.3816	4.1709
Within Groups	826,973.9	30	27,565.795	…	…	…
Total	848,710	31	…	…	…	…

Table 5 Loading conditions from the pilot at different rotorcraft configuration

Rotorcraft Configuration	Loading Conditions (lbs)	
	Pilot A	**Pilot B**
Clean	1,221	1,293
	1,293	1,257
	1,185	1,221
	1,221	1,221
	1,293	1,257
	1,400	1,329
	1,329	1,257
	1,329	1,293
ESSS	1,479	1,516
	1,641	1,623
	1,572	1,406
	1,465	1,703
	1,621	1,292
	1,720	1,340
	1,468	1,625
	1,676	1,446

Table 6 ANOVA 2 factor summary for rotorcraft configuration and pilot

Summary	**Pilot A**	**Pilot B**	**Total**
Clean Configuration			
Count	8	8	16
Sum (lbs)	10,271	10,128	20,399
Average (lbs)	1,283.875	1,266	1,274.938
Variance	5,058.125	1,388.571	3,093.663
ESSS Configuration			
Count	8	8	16
Sum (lbs)	1,2642	11,951	24,593
Average (lbs)	1,580.25	1,493.875	1,537.063
Variance	10,035.93	21,790.7	1,6841.93
Total			
Count	16	16	…
Sum (lbs)	2,2913	22,079	…
Average (lbs)	1,432.063	1,379.938	…
Variance	30,467.4	24,664.2	…

Table 7 ANOVA 2 factor output for rotorcraft configuration interaction with pilot

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	549,676.1	1	549,676.1	57.44744	2.95E-08	4.195972
Columns	21,736.13	1	21,736.13	2.271674	0.142958	4.195972
Interaction	9,384.5	1	9,384.5	0.980788	0.330486	4.195972
Within	267,913.3	28	9,568.33	…	…	…
Total	848,710	31	…	…	…	…

Table 8 shows the loads for each pilot with respect to the 2 different bank angles as they completed the different flight maneuvers on the flight card. These data were then summarized by finding the sum, average, and variance of each pilot with respect to the rotorcraft configuration shown in Table 9. Table 10 shows the ANOVA output as it is printed from Excel in which the null hypothesis is accepted since the p-values are greater than 0.05. The final factor interaction with the pilot is show in Table 11, which are the loads for each pilot with respect to the 2 different airspeeds. Table 12 summarizes the data by finding the sum, average, and variance of each pilot with respect to the rotorcraft speed. Table 13 shows the ANOVA output as it is printed from Excel in which the null hypothesis is accepted since the p-values are greater than 0.05.

As for the pilot seating arrangement, since there was no pilot variability found, an assumption was made that there is no variability between pilot/copilot seats. Although the pilot seating arrangement was not recorded, the CERDEC chief test engineer confirmed that the pilots stayed in the same seats for all of the testing because the pilot and copilot were known for all of the flights. Additional data was collected with different pilots, not in the formal test set. These results were also analyzed against the formal test set, and no difference in pilot seating was observed—further supporting the results obtained.

Table 8 Loading conditions from the pilot at different bank angles

Angle of Bank (deg°)	Loading Conditions (lbs)	
	Pilot A	**Pilot B**
35	1,221	1,293
	1,185	1,221
	1,293	1,257
	1,329	1,257
	1,479	1,516
	1,572	1,406
	1,621	1,292
	1,468	1,625
45	1,293	1,257
	1,221	1,221
	1,400	1,329
	1,329	1,293
	1,641	1,623
	1,465	1,703
	1,720	1,340
	1,676	1,446

Table 9 ANOVA 2 factor summary for angle of bank and pilot

Summary	**Pilot A**	**Pilot B**	**Total**
35˚			
Count	8	8	16
Sum (lbs)	11,168	10,867	22,035
Average (lbs)	1,396	1,358.375	1,377.188
Variance	26,274	20,901.13	22,392.56
45˚			
Count	8	8	16
Sum (lbs)	11,745	11,212	22,957
Average (lbs)	1,468.125	1,401.5	1,434.813
Variance	36,040.7	30,888	32,417.1
Total			
Count	16	16	…
Sum (lbs)	22,913	22,079	…
Average (lbs)	1,432.063	1,379.938	…
Variance	30,467.4	24,664.2	…

Table 10 ANOVA 2 factor output for angle of bank interaction with pilot

Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Sample	26,565.13	1	26,565.13	0.931262	0.342797	4.195972
Columns	21,736.13	1	21,736.13	0.761977	0.390135	4.195972
Interaction	1,682	1	1,682	0.058964	0.809911	4.195972
Within	798,726.8	28	28,525.96	…	…	…
Total	848,710	31	…	…	…	…

Table 11 Loading conditions from the pilot at different speeds

Speed (kts)	Loading Conditions (lbs)	
	Pilot A	Pilot B
80	1,221	1,293
	1,293	1,257
	1,185	1,221
	1,221	1,221
	1,479	1,516
	1,641	1,623
	1,572	1,406
	1,465	1,703
100	1,293	1,257
	1,400	1,329
	1,329	1,257
	1,329	1,293
	1,621	1,292
	1,720	1,340
	1,468	1,625
	1,676	1,446

Table 12 ANOVA 2 factor summary for speed and pilot

Summary	Pilot A	Pilot B	Total
80 kts			
Count	8	8	16
Sum (lbs)	11,077	11,240	22,317
Average (lbs)	1,384.625	1,405	1,394.813
Variance	31,142.27	35,830	31,364.43
100 kts			
Count	8	8	16
Sum (lbs)	11,836	10,839	22,675
Average (lbs)	1,479.5	1,354.875	1,417.188
Variance	29,001.43	15,586.13	24,949.23
Total			
Count	16	16	…
Sum (lbs)	22,913	22,079	…
Average (lbs)	1,432.063	1,379.938	…
Variance	30,467.4	24,664.2	…

Table 13 ANOVA 2 factor output for speed interaction with pilot

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	4,005.125	1	4,005.125	0.143605	0.707582	4.195972
Columns	21,736.13	1	21,736.13	0.779353	0.384856	4.195972
Interaction	42,050	1	42,050	1.507711	0.229711	4.195972
Within	780,918.8	28	27,889.96	…	…	…
Total	848,710	31	…	…	…	…

6. Conclusions

For the null hypothesis regarding pilot variability, the evidence shows that there is no change in effectiveness at the 95% confidence level. This was proven since none of the test analyses showed any justifiable pilot variability with respect to the 95th percentile. This was enough information to accept the null hypothesis and as such state that pilot variability has no effect on the accuracy of the regime recognition algorithms. For the case of the loads, seen by the swash plate, for each pilot with respect to the 2 different rotorcraft weight configurations, there was variability found between the clean configuration and the ESSS. This was not pilot variability, since the variability in the sampling was caused by a large difference between having an empty rotorcraft and a rotorcraft that is weighed down with additional stores. Therefore, despite a p-value for the sample smaller than 0.05, it is a p-value that represents variability different than pilot variability.

ARL built a DOE that subjectively assessed the variability with regards to the pilots to remove any bias that might be caused by using different pilots. The method reduced the amount of flight time and number of pilots by approximately five-ninths and one-third, respectively, and minimized the effect that challenges of this test had on the results. The fact that there is no pilot variability in the current process of regime recognition for the current HUMS software on the UH-60 rotorcraft means that there is no bias in the software with respect to changing pilots.

7. References

1. ADS-79D-HDBK. Aeronautical design standard handbook for condition based maintenance systems for US army aircraft systems. Revision D. Redstone Arsenal (AL): Army Aviation and Missile Reseach, Development, and Engineering Command (US), Aviation Engineering Directorate; 2013 Mar 7.

2. Oehlert GW. A first course in design and analysis of experiments. New York (NY): WH Freeman; 2000.

3. FAA advisory circular. Advisory circular 29 miscellaneous guidance 15. 2003.

List of Symbols, Abbreviations, and Acronyms

ADS-79D	Aeronautical Design Standard Handbook for Condition Based Maintenance Systems for US Army Aircraft
ANOVA	Analysis of Variance
ARL	US Army Research Laboratory
CERDEC	US Army Communications-Electronics Research, Development and Engineering Center
df	degrees of freedom
DOE	Design of Experiment
ESSS	External Stores Support System
FRR	flight regime recognition
HUMS	Health and Usage Maintenance Systems
MS	Mean Squares
SS	sum of square

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIRECTOR
(PDF) US ARMY RESEARCH LAB
RDRL CIO LL
IMAL HRA MAIL & RECORDS
MGMT

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 DIR USARL
(PDF) RDRL VTM
N BRADLEY