

# AFRL-RX-WP-TP-2012-0411

# EXPERIMENTAL RESOURCE ALLOCATION FOR STATISTICAL SIMULATION OF FRETTING FATIGUE PROBLEM (PREPRINT)

Patrick Golden Metals Branch Structural Materials Division

Harry R. Millwater, Carolina Dubinsky, and Gulshan Singh University of Texas at San Antonio

AUGUST 2012 Interim

Approved for public release; distribution unlimited.

See additional restrictions described on inside pages

**STINFO COPY** 

## AIR FORCE RESEARCH LABORATORY MATERIALS AND MANUFACTURING DIRECTORATE WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7750 AIR FORCE MATERIEL COMMAND UNITED STATES AIR FORCE

REPORT DOCUMENTATION PAGE						Form Approved OMB No. 0704-0188		
The public reporting bu sources, gathering and information, including s Davis Highway, Suite 1 collection of informatio	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, searching existing data accepted 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 this 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. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS</b> .							
1. REPORT DAT	ГЕ <i>(DD-MM-YY</i> )	)	2. REPORT TYPE			3. DATE	S COVERED (From - To)	
August 20	)12		Technical Par	ber		1 Jul	v 2012 – 1 August 2012	
4. TITLE AND S	UBTITLE						5a. CONTRACT NUMBER	
EXPERI	MENTAL RE	SOURCE AL	LOCATION FO	R STATIST	ICAL		In-house	
	TION OF FR	ETTING FA	FIGUE PROBLE	M (PREPRI	NT)		5b. GRANT NUMBER	
SINICLA			TIOULTRODLE		111)		5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)							5d. PROJECT NUMBER	
Detrick G	oldon (AEDI	( <b>PYCM</b> )					/3/9	
I attick O	Millwoton C	MACIVI)	ury and Culahan	Sinch (Unit	omity of T	waa at	5e TASK NUMBER	
Harry K.	Millwater, Ca	aronna Dubins	sky, and Guisnan	Singn (Univ	ersity of Te	exas at		
San Anto	nio)							
							SI. WORK UNIT NUMBER	
							X05W	
7. PERFORMIN	G ORGANIZATIO	ON NAME(S) AN	D ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER	
Metals Bra	anch						AFRL-RX-WP-TP-2012-0411	
Structural	Materials Divi	sion						
2230 Tent	h Street							
Wright-Pa	tterson AFB, C	DH 45433-7750						
9. SPONSORIN	G/MONITORING		E(S) AND ADDRESS	S(ES)			10. SPONSORING/MONITORING	
Air Force	Research La	boratory					AGENCT ACKONTM(S)	
Materials	and Manufac	cturing Directo	orate				AFRL/RXCM	
Wright-Patterson Air Force Base, OH 45433-7750					11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)			
Air Force	Materiel Co	mmand, Unite	d States Air Ford	ce			AFRL-RX-WP-TP-2012-0411	
12. DISTRIBUT	ON/AVAILABIL	ITY STATEMENT						
Approved	l for public re	elease; distribu	tion unlimited. F	Preprint to be	submitted	to AIAA	A SDM Conference.	
13. SUPPLEME	NTARY NOTES							
This wor	k was funded	in whole or ir	part by Departn	nent of the A	ir Force In	house. T	The U.S. Government has for itself	
and other	s acting on its	s behalf an un	imited, paid-up.	nonexclusiv	e, irrevocab	le world	lwide license to use, modify.	
reproduce	e, release, per	form, display.	or disclose the y	vork by or o	behalf of t	he U.S.	Government, PA Case Number	
and clear	ance date: 88.	ABW-2012-2	173, 12 April 201	12. This doc	ument conta	ains colo	or.	
14 ABSTRACT								
Estimatio	n of statistica	l moments fro	om simulation i	e., mean and	standard de	viation	of an output, may involve large	
uncertain	ty caused by	the variability	in the input rand	lom variable	The alloc	ation of	resources to obtain more	
experime	ntal data can	reduce the var	iance of the outr	out momente	(mean and	standard	deviation) The methodology	
proposed	and executed	lused an optir	nization method	to determine	the optimal	l numbe	r of additional experiments required	
to minim	ize the varian	ce of the outp	ut moments give	n a constrain	$\Delta = \frac{1}{2} + $	to gen	erate the output moments based on	
the mome	ants of the ine	ut variables	us includes give	The mothod	luced the w	10 gent	ate t_distribution and the Wishart	
diataihurt	on to concrete	realizations	of the population	moon and -	nulotion of	women	a of the input verichlas	
	on to generate	bod was sufficient	iont to handled in	ndenon dent	putation Co	od voria	t of the input variables,	
respective	to minimize t	hou was suffic	for to failure	mean and a	tondord day	eu varia	oles. A neunig laugue problem was	
			cycles-to failule	e mean and s		Tation.		
fretting f	EKMS atigue: resour	ce allocation.	uncertainty auan	tification: co	onfidence in	tervals		
		N OF:						
TO. SECURITY			OF	PAGES	ISA. NAME	or RESP	don	
a. REPORT	Unclassified	Unclassified	ABSTRACT:	18	Patri		UCII LIMBER (Include Area Code)	
Unclassified	Cherassified	Cherassified	SAR		ISD. IELEI N/A			
					1 N/ PA			

Г

Standard Form 298 (Rev. 8-98) Prescribed by ANSI Std. Z39-18

## Experimental Resource Allocation for Statistical Simulation of Fretting Fatigue Problem

Carolina Dubinsky<sup>1</sup>, Gulshan Singh<sup>2</sup>, and Harry R. Millwater<sup>3</sup> University of Texas at San Antonio, TX 78249, USA

Patrick Golden<sup>4</sup> Air Force Research Laboratory, Wright-Patterson AFB, OH 45433, USA

### Abstract

Estimation of statistical moments from simulation, i.e., mean and standard deviation of an output, may involve large uncertainty caused by the variability in the input random variables. The allocation of resources to obtain more experimental data can reduce the variance of the output moments (mean and standard deviation). The methodology proposed and executed used an optimization method to determine the optimal number of additional experiments required to minimize the variance of the output moments given a constraint. A method to generate the output moments based on the moments of the input variables was implemented. The method used the multivariate t-distribution and the Wishart distribution to generate realizations of the population mean and population covariance of the input variables, respectively. This method was sufficient to handled independent and correlated variables. A fretting fatigue problem was explored to minimize the variance of cycles-tofailure mean and standard deviation. The optimal number of additional experiments required for each random variable depended on the number of initial data points, the influence of the variable in the output function, the cost of each additional experiment and the variance of the sample mean.

## Nomenclature

$A_i$	=	Constants in the output function for variable $X_i$
b	=	Funds available for additional experiments
$C_{G_i}$	=	Cost of each additional experiment for group $G_i$
$C_{X_i}$	=	Cost of each additional experiment for variable $X_i$
$D_{G_i}$	=	Number of additional experiments for group $G_i$
$D_{X_i}$	=	Number of additional experiments for variable $X_i$
$E_{G_i}$	=	Number of initial data points for group $G_i$
$E_{X_i}$	=	Number of initial data points for variable $X_i$
$G_i$	=	Group of random variables $X$
$G_{D_i}$	=	Values of additional experiments for group $G_i$
$G_{E_i}$	=	Values of initial data points for group $G_i$
gbest.	$\chi_{=}$	Best position encountered by any particle in PSO (Global best)
k	=	Optimization iteration number
MCS	=	Monte Carlo sampling
$N_{f}$	=	Cycles-to-failure
$n_{X}$	=	Number of total data points for variable $X_i$

<sup>&</sup>lt;sup>1</sup> Graduate Research Assistant, Department of Mechanical Engineering and AIAA Member

<sup>&</sup>lt;sup>2</sup> Post-Doctoral Researcher, Department of Mechanical Engineering and AIAA Member

<sup>&</sup>lt;sup>3</sup> Professor, Department of Mechanical Engineering and AIAA Member

<sup>&</sup>lt;sup>4</sup> Materials Research Engineer, Materials and Manufacturing Directorate and AIAA Member

$N_p$	=	Multivariate normal distribution
p	=	Number of variables in the output function
pbest.	$x_i =$	Best position found by $i^{th}$ particle in all previous iterations in PSO (Personal best)
PDF	=	Probability density function
$r_{SU}$	=	Particle swarm optimization
$q_1$	_	Social weight in PSO
$Y_2 r_1, r_2$	=	Number from uniform distribution between 0 and 1 for the $i^{th}$ PSO particle
$S_{X_i}$	=	Sample standard deviation for variable $X_i$
$S_{X_{i}}^{2}$	=	Sample variance for variable $X_i$
$t_{n_{X_i}} - 1$	=	Univariate Student's t-distribution with <i>n-1</i> degrees of freedom
$t_{p,n_{X_i}}$ -1	=	Multivariate t-distribution of $p$ variables with $n-1$ degrees of freedom
$v_i^k$	=	Velocity of $i^{th}$ PSO particle at iteration $k$
$v_i^{k+1}$	=	Velocity of $i^{th}$ PSO particle at iteration $k+1$
W	=	Inertia weight in PSO
$W_p$	=	Wishart distribution
$X_{D_i}$	=	Values of $X_i$ for additional experiments $D_{X_i}$
$X_{E_i}$	=	Values of $X_i$ for initial data points $E_{X_i}$
$X_i$	=	Input variable $i^{th}$ in the output function
$\underline{X}_{i}^{(j)}$	=	Observation $j^{th}$ of the $X_i$ input variable
$X_{X_i}$	=	Sample mean of variable $X_i$
$x_i^k$	=	Position of $i^{th}$ PSO particle at iteration $k$
Ż	=	Output function (or response function)
<u>Greek L</u> 2	etters	
χ	=	Chi-square distribution
$\mu_{X_i}$	=	Input population mean of variable $X_i$
$\mu_{Z}$	=	Output mean
$ ho_{ij}$	=	Correlation coefficient between $X_i$ and $X_j$
Σ	=	Covariance matrix
$\sigma_{X_i}$	=	Input population standard deviation of variable $X_i$
$\sigma_{X_i}$	=	Input population variance of variable $X_i$
$\sigma_{\!\scriptscriptstyle\mu_Z}$	=	Standard deviation of the output mean
$\sigma_{\sigma_Z}$	=	Standard deviation of the output standard deviation
$\sigma^{ m ong}_{\mu_Z}$	=	Standard deviation of the output mean based on the original data
$\sigma^{\scriptscriptstyle opt}_{_{\mu_{Z_{.}}}}$	=	Standard deviation of the output mean based on the optimum solution
$\sigma^{ m orig}_{\sigma_{ m z}}$	=	Standard deviation of the output std. dev. based on the original data
$\sigma^{\scriptscriptstyle opt}_{\!\sigma_z}$	=	Standard deviation of the output std. dev. based on the optimum solution
$\sigma_{Z}$	=	Output standard deviation
$\sigma_{\!Z}^{\!ar{2}}$	=	Output variance

 $\zeta$  = Sample covariance matrix  $\Omega$  = Data set

## I. Introduction

THE presence of uncertainty in risk and reliability analysis is unavoidable; it is an important part of the planning, executing, and decision-making process. To develop estimates, researchers must rely on available data that is often limited and contains variability. Moreover, they have to rely on estimation or predictions based on idealized models that involve additional uncertainty compared to reality [1]. Statistical estimates from simulation, such as mean and standard deviation of the output or the probability-of-failure (probability of exceeding a limit), often involve significant uncertainty caused by the variability in the input random variables. The probability distribution of the input variables may be developed from the limited available data, thus the sample mean and standard deviation of the input are also random variables dependent upon the sample size. The resulting uncertainty in the estimated output moments can be significant and should be taken into account when making decisions [2] [3] [8] [10] [11] [13] [14].

Previous research has focused on the quantification of uncertainty caused by the variability of the input random variables using confidence intervals of the output model. Numerous authors have developed methods to calculate these confidence intervals. On the other hand, less work has been done to reduce the uncertainty of the output model, and a very limited number of authors have tried to increase the confidence of the output model by allocating resources to obtain more experimental data of the input random variables.

Several methods have been proposed to accomplish the computation of the confidence intervals of the statistical estimates (the output moments or the probability-of-failure). Most authors have developed methods to estimate the probability-of-failure using a first-order reliability method (FORM) [10][11][8][14]. FORM estimates the shortest distance, known as reliability index, from the origin of a standard normal variable space to a design point (most probable point) on a limit state. The limit state is the boundary between the safe and unsafe region [11]. The uncertainty present in the distribution of the input parameters is quantified by obtaining the confidence intervals of the reliability index (or safety index).

The methods to calculate the confidence intervals of the safety index in reliability analysis are accepted for many problems if the most probable point can be located, and the limit state function (boundary between the safe and unsafe regions) can be approximated with a surface of first- or second-order. A more general method to obtain the reliability or probability-of-failure is using Monte Carlo simulation (MCS), and some authors have done rigorous studies on this matter [4] [7].

The most recognized strategy to determine the influence of the input parameter variation on the output model is accomplished by nesting a loop of a single output calculation within a loop that accounts for the uncertainty of the input parameter. The loop where the output moments are computed is often referred to as the "inner-loop," and the loop where the variation of the input parameter is taken into account is usually referred to as the "outer-loop." The limitation of this method is the high computational cost occasioned by the nested loops. The accomplishment of the nested simulation can represent a non-trivial problem, and several authors have developed computational strategies to address this issue. The common strategy involves using a surrogate model, such as a response surface, to approximate the probability-of-failure as a function of the input moments [2][3]. However, the accuracy of the statistical estimates depends on the quality of the surrogate model.

Several authors have studied the variation of statistical estimates from simulation, such as mean and standard deviation of the output or the probability-of-failure, in the presence of uncertainty in the input parameters. They have tried to quantify the variation and delimit the reliability with confidence intervals. Most of the authors have agreed and developed strategies to address the computational complexity by using surrogate models to replace the inner-loop in the nested reliability analysis; however, very limited research has been done to increase the confidence in the output model by taking any actions over the input parameters, such as allocating resources to obtain more experimental data of the input random variables.

Urbina et al. [13] implemented a hierarchical approach to minimize the mean and the range of the probability-offailure by allocating resources to obtain additional experimental data of the input variables. The input parameters were obtained from an empirical cumulative distribution function developed from the observed data. These parameters were introduced into a Bayesian network to obtain the system response. The system response was compared to an expected performance measure to calculate the probability-of-failure. A multi-objective optimization problem was solved using a grid search approach and the constraint was a function of cost of the additional experimental data. The main purpose of this research was to develop a methodology to reduce the variation of an output moment using an optimization algorithm that varied the number of data points obtained for each input variable. Another purpose was to develop a method to generate realization of the population mean and population covariance of the input random variables. The method required to handle independent and correlated variables. The optimal allocation methodology aimed to find the optimal additional experimental data needed to better characterize the moments of the input probability density functions (PDFs) in order to minimize the variance of the output moments, such as mean and standard deviation, subject to a constraint. The methodology combined a single-objective optimization algorithm with a nested-loop arrangement. The output moments were calculated analytically for efficiency; however the methodology is not limited to such models.

## **II. Simulation of Statistical Moments**

The sample mean,  $\overline{X}_{X_i}$ , and sample variance,  $S_{X_i}^2$ , become random variables when sampled multiple times from the same population. A probabilistic distribution of the sample moments may be obtained from these multiple samples. These distributions are known as sampling distributions [15] and are used to simulate the input population mean and standard deviation.

#### A. Population mean

In general, the standard deviation of the population,  $\sigma_{X_i}$ , is unknown and needs to be estimated with the sample standard deviation,  $S_{X_i}$ . As a result, the random variable  $(\overline{X}_{X_i} - \mu_{X_i})/(S_{X_i}/\sqrt{n_{X_i}})$  follows a t-distribution as

$$\frac{\overline{X}_{X_i} - \mu_{X_i}}{S_{X_i} / \sqrt{n_{X_i}}} \sim t_{n_{X_i} - 1}$$

$$\tag{1}$$

where  $t_{n_{X_i}-1}$  is the Student's t-distribution with  $n_{X_i} - 1$  degrees of freedom [1]. Consequently, realizations of the population mean,  $\mu_{X_i}$ , can be determined as

$$\mu_{X_{i}} = \overline{X}_{X_{i}} - t_{n_{X_{i}}-1} \frac{S_{X_{i}}}{\sqrt{n_{X_{i}}}}$$
<sup>(2)</sup>

In the case of correlated variables, the procedure is to generate realizations of the multivariate t-distribution and compute the population mean,  $\mu_{x_i}$ , as follows

$$\mu_{X_{i}} = \overline{X}_{X_{i}} - t_{p_{i}, n_{X_{i}}-1} \frac{S_{X_{i}}}{\sqrt{n_{X_{i}}}}$$
(3)

where  $\overline{X}_{X_i}$  is the sample mean of variable  $X_i$ ,  $S_{X_i}$  is the sample standard deviation,  $n_{X_i}$  is the number of observations, and  $t_{p_i,n_{X_i}-1}$  is the *i*<sup>th</sup> realization from the *p*-variate t-distribution with  $n_{X_i} - 1$  degrees of freedom, location vector zero and scale matrix  $\zeta$ . In this approach, it is assumed that the number of data points is the same for all random variables that are correlated. The sample covariance,  $\zeta$ , is a *p x p* matrix calculated as follows

$$\zeta = \frac{1}{n-1} \sum_{j=1}^{n} (X^{(j)} - \overline{X}) (X^{(j)} - \overline{X})'$$
(4)

where  $X_i^{(j)}$  is the  $j^{th}$  observation of the  $X_i$  input variable with i = 1, ..., p and j = 1, ..., n. The sample covariance,  $\zeta$ , is the unbiased estimator of the covariance,  $\Sigma$ .

## **B.** Population Variance and Covariance Matrix

Similarly, using the probability distribution of the sample variance,  $S_{X_i}^2$ , the random variable  $(n_{X_i} - 1)S_{X_i}^2 / \sigma_{X_i}^2$  has the following distribution

$$\frac{(n_{X_i} - 1)S_{X_i}^2}{\sigma_{X_i}^2} \sim \chi^2_{n_{X_i} - 1}$$
(5)

where  $\chi^2_{n_{\chi_i}-1}$  is the chi-square distribution with  $n_{\chi_i} - 1$  degrees of freedom [1]. As a consequence, realizations of the population variance,  $\sigma^2_{\chi_i}$ , can be determined as follows

$$\sigma_{X_i}^2 = \frac{(n_{X_i} - 1)S_{X_i}^2}{\chi_{n_{X_i} - 1}^2} \tag{6}$$

In the case of correlated variables, the input population covariance matrix was obtained using a Bayesian approach<sup>5</sup>. A commonly used prior distribution is

$$p(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{p+1}{2}}}$$
(7)

where  $|\Sigma|$  is the determinant of the population covariance matrix  $\Sigma$ . Considering  $\Omega = (X^{(1)}, X^{(2)}, ..., X^{(n)})$  to be the observed data, then the likelihood  $p(\Omega|\mu,\Sigma)$ , and prior distribution  $p(\mu,\Sigma)$ , determine the joint-distribution  $p(\Omega,\mu,\Sigma)$ . From this, the posterior distribution of the unknown parameters  $p(\mu,\Sigma|\Omega)$  can be obtained. The Bayesian approach prescribes the use of the posterior distribution to make inference about unknown parameters, so in particular it can be used to simulate values for the unknown means, and covariance. From the above models the inverse of the population covariance matrix,  $\Sigma^{-1}$ , conditioned on the data  $\Omega$  has the following distribution

$$(\Sigma^{-1} \mid \Omega) \sim W_p \left( \left( (n_{X_i} - 1)\xi \right)^{-1}, n_{X_i} - 1 \right)$$
(8)

where  $W_p\left(\left((n_{X_i}-1)\zeta\right)^{-1}, n_{X_i}-1\right)$  represents the Wishart distribution with  $n_{X_i}-1$  degrees of freedom, and scale matrix  $\left((n_{X_i}-1)\zeta\right)^{-1}$ . To obtain the population covariance matrix,  $\Sigma$ , it is necessary to sample from the Wishart distribution given by Eq. (8) and invert the values, thus yielding

$$\Sigma^{-1} = W_p \left( \left( (n_{X_i} - 1)\zeta \right)^{-1}, n_{X_i} - 1 \right)$$
(9)

## **III. Optimal Allocation of Resources**

Different approaches have been studied for quantifying the uncertainty in the statistical estimates, such as mean and standard deviation of the output or the probability-of-failure, caused by the variation of the input random variables. To date, there has been little development on how to reduce the variation of the output moment distribution by taking action over the input variables. The action considered in this work was to add additional data or experiments to the input variables to better characterize the mean and standard deviation of the input probability density function (PDFs). The methodology aims to determine the optimal number of experiments required to minimize the variance of the output moments given a constraint. The methodology can also be defined, as what experiments should be conducted in order to improve the confidence in the output moments of a probabilistic problem.

## A. Methodology

A schematic of the computational approach is shown in Figure 1. The methodology proposed is general and can be applied in any field where a reduction of the variance of a statistical estimate is required.

<sup>&</sup>lt;sup>5</sup> Personal communication with Dr. Victor De Oliveira, Associate Professor at the Department of Management Science and Statistics at the University of Texas at San Antonio, Texas



Figure 1. Schematic flow chart of optimal allocation methodology

In practice, the determination of the distribution of the output moments are often computed using Monte Carlo sampling. The sampling requires an iterative repetition of actions called a "loop." This loop is often within another cycle of actions, therefore, it is called the "inner-loop." The inner loop generates realizations of the output moments, such as mean or standard deviation. The "outer-loop" determines the distribution and standard deviation of the output moments. The approach considered in this research was to use an optimization model combined with the nested-loop arrangement to minimize the standard deviation of the output moments.

Every computation of the nested-loop is an iteration of the optimization process. In every iteration, random numbers of additional experimental data are tested; the outcome of each iteration is the lowest value of the standard deviation of the output moment. The optimization process is repeated until the number of iterations is reached. The final result of the optimization is the optimal additional experiments that returned the minimum value of the standard deviation of the output moment. This method is a single-objective optimization; only the standard deviation of one output moment, such as output mean or output standard deviation, can be optimized at a time.

The constraint of the optimization model is  $\sum_{i=1}^{p} C_{X_i} D_{X_i} \le b$ . Where  $C_{X_i}$  is the cost of each additional experiment,  $D_{X_i}$  is the number of additional experiments of variable  $X_i$ , and b is the total funds available. The statistical process to minimize the standard deviation of the output moment subject to the constraint is explained as follows:

- 1. Initial data of input variable  $X_{E_i}$  are provided
- 2. Additional experimental data,  $X_{D}$ , is generated
- 3. The input sample mean,  $\overline{X}_{X_i}$ , and input sample covariance,  $\zeta$ , are calculated
- 4. Realizations of the *p*-variate t-distribution,  $t_p$ , and Wishart distribution,  $W_p$ , are used to simulate the population mean,  $\mu_{X_i}$ , and population covariance,  $\Sigma$ , as shown in Eq. (3) and Eq. (9), respectively ("outer-loop")
- 5. According to the objective, the output mean,  $\mu_z$ , or output standard deviation,  $\sigma_z$ , is calculated ("inner-loop")

- 6. Steps 3 and 4 are repeated to generate a distribution of the output moment
- 7. The standard deviation of the output moment is calculated
- 8. The optimization algorithm varies the number of additional experimental data (Step 2) and determines the optimal number of additional experiments,  $D_{x}$ .

#### **B.** Optimization Method

The optimization problem of determining the subsequent experiments needed to reduce the variance of the output moment subject to a cost constraint is formulated as follows:

Objective :  $Minimize(\sigma_{\mu_z}, \sigma_{\sigma_z})$ Constraint :  $\sum_{i=1}^{p} C_{X_i} D_{X_i} \le b$ Variable Bounds :  $D_{X_i}^{Lower} < D_{X_i} < D_{X_i}^{Upper}$ 

where  $\sigma_{\mu_z}$  is the standard deviation of the output mean,  $\sigma_{\sigma_z}$  is the standard deviation of the output standard deviation,  $C_{X_i}$  is the cost of each additional experiment,  $D_{X_i}$  is the number of additional experiments, b are the funds available, and  $D_{X_i}^{Lower}$ ,  $D_{X_i}^{Upper}$  are the lower and upper bounds of the additional experiments  $D_{X_i}$ , respectively.  $D_{X_i}^{Upper}$  is obtained by dividing the total funds available by the cost of each additional experiment, and  $D_{X_i}^{Lower}$  is zero. This optimization is single-objective; therefore, only the standard deviation of the output mean,  $\sigma_{\mu_z}$ , or the standard deviation of the output standard deviation,  $\sigma_{\sigma_z}$ , is minimized.

The optimization of a non-linear function of integer variables and the high-computational cost associated with a function evaluation suggests that a population-based approach is suitable to solve the problem. A particle swarm optimization (PSO) was selected because of the ease of implementation and the lower user parameters.

Particle swarm optimization (PSO) is a population-based method used in the optimization of non-linear functions; it was proposed in 1995 by Kennedy and Eberhart [6]. PSO is a swarm intelligence method that models social behavior of a population (swarm) of agents (particles) interacting to find a simulated target on a search space. In the particle swarm optimization process, the velocity and position of each particle is iteratively adjusted as shown in Eq. (10) and Eq. (11), respectively

$$v_i^{k+1} = wv_i^k + q_1 r_{il} (pbestx_i - x_i^k) + q_2 r_{i2} (gbestx - x_i^k)$$
(10)

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{11}$$

The velocity is defined as a change in magnitude of the design variable from one iteration to another, and the position is described as the design variable unit, in this research, as the number of additional experimental data  $D_{x_i}$ . The particles move according to a communication structure thought of as a social network. At iteration k the velocity of the  $i^{th}$  particle  $v_i^k$  is updated according to its own current velocity value, the best position encountered by the  $i^{th}$  particle in all previous iterations (particle best,  $pbestx_i$ ), the best position encountered by any particle so far (global best, gbestx), and the inertia weight, w, that controls the impact of the previous velocity. The particles are attracted toward the positions of  $pbestx_i$  and gbestx; the strength of the attraction is controlled by  $q_1$  (individual weight) and  $q_2$  (social weight). Randomness is introduced for good space exploration via  $r_{i1}$  and  $r_{i2}$  which are random numbers from a uniform distribution on the interval between 0 and 1. The position of the particle is updated using its current position value  $x_i^k$  and the newly computed velocity  $v_i^{k+1}$  [12]. The constants w,  $q_1$ , and  $q_2$ , are empirical. Trelea et al. [12] have conducted several experiments with different combinations of these constants recommended by other authors and concluded that the best results published are

Inertia weight (w): 0.729 Cognitive constant  $(q_1)$ : 1.494 Social constant  $(q_2)$ : 1.494

7

The number of particles and the number of iterations were selected according to the complexity of the problem. The constraint used in the method was controlled by using a penalty function, where a significant penalty was assigned to the objective function value if the constraint was exceeded. This strategy forced the particles to move towards the feasible design space where the target was located.

## **IV. Fretting Fatigue**

The optimal allocation methodology was applied to a fretting fatigue problem. Fretting is the wear damage caused when a material is compressed against one another in the presence of oscillatory displacements. The wear and high local stresses cause nucleation of cracks that reduce the fatigue life. Fretting fatigue is a major problem in the aerospace industry. The damage that occurs from fretting fatigue causes structural failure that may be very costly. Previous research of fretting fatigue has been done by Golden et al. [5] who performed a probabilistic fretting fatigue life prediction analysis of Ti-6Al-4V dovetail specimens.

The statistical data given in reference [5] was used in this case study. The fretting fatigue problem consisted of 20 random variables with mean, standard deviation, and correlation values determined from experimental data. The statistics of the random variables are shown in Table 1.

Table 1. Random Variables Statistics							
Random Variable	Variable	Mean,	St. dev.,	Distribution Type			
Kanuoni variabie	No.	$\mu_{X_i}$	$\sigma_{\!\scriptscriptstyle X_i}$	Distribution Type			
Initial Crack	$X_1$	15.1	8.48	Lognormal			
Friction Coeff.	$X_2$	0.302	0.021	Correlated Normal			
Partial Slip Slope	$X_3$	1.96	0.12	$ \rho_{23} = -0.375 $			
Crack Growth	$X_4$	-14.6	0.486	Correlated Normal			
Crack Growth	$X_5$	7.19	0.715	$ \rho_{45} = -0.9973 $			
Crack Growth	$X_{6}$	-11.8	0.157	Correlated Normal			
Crack Growth	$X_7$	3.81	0.146	$\rho_{_{67}} = -0.9751$			
Pad Profile	$X_8$	0.181	5.84E-03				
Pad Profile	$X_9$	-2335	410				
Pad Profile	$X_{10}$	2333	411				
Pad Profile	$X_{11}$	-1612	37.7				
Pad Profile	$X_{12}$	2289	379				
Pad Profile	$X_{13}$	1620	35.4				
Pad Profile	$X_{14}$	-0.183	4.96E-03	Correlated Normal O(see Table 2)			
Pad Profile	$X_{15}$	-2.00E-04	6.20E-04	<i>p</i> (			
Pad Profile	$X_{16}$	-1.21E-06	1.01E-06				
Pad Profile	$X_{17}$	1.53E-10	6.38E-10				
Pad Profile	$X_{18}$	9.80E-13	6.16E-13				
Pad Profile	$X_{19}$	-3.77E-17	1.87E-16				
Pad Profile	$X_{20}$	-3.80E-19	1.59E-19				

**Table 2. Pad Profile Correlation Coefficients** 

	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	<i>X</i> <sub>13</sub>	$X_{14}$	$X_{15}$	$X_{16}$	<i>X</i> <sub>17</sub>	<i>X</i> <sub>18</sub>	$X_{19}$	$X_{20}$
$X_8$	1	-0.079	0.078	-0.119	-0.051	0.184	-0.364	0.153	0.175	-0.107	-0.130	0.073	0.112
$X_9$	-0.079	1	-1.000	-0.921	-0.404	0.321	-0.122	0.093	-0.205	-0.203	0.281	0.275	-0.246
$X_{10}$	0.078	-1.000	1	0.921	0.404	-0.321	0.121	-0.092	0.207	0.203	-0.282	-0.275	0.247
$X_{11}$	-0.119	-0.921	0.921	1	0.325	-0.372	0.270	-0.019	0.104	0.136	-0.201	-0.247	0.219
$X_{12}$	-0.051	-0.404	0.404	0.325	1	-0.899	-0.204	-0.102	0.079	0.065	-0.116	0.002	0.123
$X_{13}$	0.184	0.321	-0.321	-0.372	-0.899	1	-0.033	0.010	-0.004	0.062	0.021	-0.147	-0.061
$X_{14}$	-0.364	-0.122	0.121	0.270	-0.204	-0.033	1	-0.095	-0.242	-0.017	0.142	0.078	-0.133
$X_{15}$	0.153	0.093	-0.092	-0.019	-0.102	0.010	-0.095	1	0.140	-0.876	-0.027	0.627	0.059
$X_{16}$	0.175	-0.205	0.207	0.104	0.079	-0.004	-0.242	0.140	1	-0.099	-0.869	0.057	0.765
$X_{17}$	-0.107	-0.203	0.203	0.136	0.065	0.062	-0.017	-0.876	-0.099	1	0.011	-0.915	-0.027
$X_{18}$	-0.130	0.281	-0.282	-0.201	-0.116	0.021	0.142	-0.027	-0.869	0.011	1	0.027	-0.944
$X_{19}$	0.073	0.275	-0.275	-0.247	0.002	-0.147	0.078	0.627	0.057	-0.915	0.027	1	-0.031
$X_{20}$	0.112	-0.246	0.247	0.219	0.123	-0.061	-0.133	0.059	0.765	-0.027	-0.944	-0.031	1

Linear regression was used to fit a predictive model of the form  $\log(N_f) = A_o + \sum A_i X_i$  to a set of 10,000 data points<sup>6</sup>, where  $N_f$  is cycles-to-failure. The linear regression coefficients  $A_o$  and  $A_i$  are shown in Table 3. The coefficient of determination  $R^2 = 0.87$ , thus about 87% of the variation of  $\log(N_f)$  is explained by the predictor variables in the model

Table 3. Linear Reg	gression Coefficients
Term	Estimate, $A_i$
Intercept, $(A_o)$	4.93
$X_1$	-4.10E+03
$X_2$	-7.48
$X_3$	2.52E-03
$X_4$	-0.18
$X_5$	-0.14
$X_6$	-0.48
$X_7$	-0.44
$X_8$	0.06
$X_9$	1.08E-04
$X_{10}$	1.05E-04
$X_{11}$	-1.95E-04

Table 3. Linear Regression Coefficients

<sup>6</sup> Provided by Dr. Patrick J. Golden from the Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson AFB, OH

$X_{12}$	-5.60E-05
$X_{13}$	-9.81E-04
$X_{14}$	0.16
$X_{15}$	8.47
$X_{16}$	1.51E+05
X <sub>17</sub>	1.05E+08
X <sub>18</sub>	4.64E+11
$X_{19}$	3.51E+14
$X_{20}$	1.26E+18
-0	

Once the predicted model,  $\log(N_f) = A_o + A_1X_1 + A_2X_2 + ... + A_{20}X_{20}$ , was validated, it was used to analytically obtain the moments of  $\log(N_f)$ . The mean and standard deviation of  $\log(N_f)$  were calculated as shown in Eq. (12) and Eq. (13) respectively

$$\mu_{\log(N_f)} = A_o + A_1 \mu_{X_1} + A_2 \mu_{X_2} + \dots + A_{20} \mu_{X_{20}}$$
(12)

and

$$\sigma_{\log(N_f)} = \sqrt{\sum_{i=1}^{20} \sum_{j=1}^{20} A_i A_j \Sigma_{ij}}$$
(13)

where  $\mu_{X_i}$  represents the mean of variable  $X_i$  and  $\Sigma_{ij}$  represents the  $i^{th}$ ,  $j^{th}$  value of the population covariance matrix,  $\Sigma$ .

Finally, the objective was to determine how many additional experiments were needed to minimize the standard deviation of  $\log(N_f)$  moments,  $(\sigma_{\mu_{\log(N_f)}}, \sigma_{\sigma_{\log(N_f)}})$ , given funds available.

The random variables were partitioned into four groups according to the correlation of the random variables. The group's distribution, cost, and initial data are shown in Table 4. The number of additional experiments required to minimize the standard deviation of  $\log(N_f)$  were determined by group.

Group	Random Variable	No.	Distribution Type	Test Cost <sup>7</sup> $C_{G_i}$	Initial Data
$G_1$	Initial Crack	<i>X</i> <sub>1</sub>	Lognormal	\$846	20
$G_2$	Friction Coeff/ Partial Slip Slope	$X_{2} - X_{3}$	Correlated Normal $\rho_{23} = -0.375$	\$4,810	17
$G_3$	Crack Growth	$X_4 - X_7$	Correlated Normal $\rho_{45} = -0.9973$ $\rho_{67} = -0.9751$	\$4,748	198
$G_4$	Pad Profile	$X_8 - X_{20}$	Correlated Normal $\rho$ (see Table 2)	\$919	77

#### Table 4. Grouping of Random Variables

## Case C-1. Minimize $\sigma_{\mu_{log(Nf)}}$

A preliminary study was performed before utilizing the optimization methodology, in which the total funds available, b = \$20,000, were allocated only for one group at a time. The maximum number of additional experiments for each group was calculated as  $D_{G_i} = \lfloor b/C_{G_i} \rfloor$ . The mean of  $\log(N_f)$  was given as

<sup>&</sup>lt;sup>7</sup> Provided by Dr. Patrick J. Golden from the Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson AFB, OH

Table 5. Reduction if total funds were spent on each group							
	Test 1	Test 2	Test 3	Test 4			
$D_{G_1}$	23	0	0	0			
$D_{G_2}$	0	4	0	0			
$D_{G_3}$	0	0	4	0			
$D_{G_4}$	0	0	0	21			
$\sigma^{ m  horig}_{\mu_{\log(Nf)}}$	0.0366	0.0366	0.0366	0.0366			
$\sigma^{ hopt}_{\mu_{\log(Nf)}}$	0.0354	0.0311	0.0365	0.0362			
% Reduction	3%	15%	0.3%	1%			
$\Sigma C_{G_i} D_{G_i}$	\$19,458	\$19,240	\$18,992	\$19,299			
Each experimental data of $G_i$ reduced $\sigma_{\mu_{log(Nf)}}$ by	0.13%	3.75%	0.08%	0.05%			

 $\mu_{\log(N_f)} = A_o + A_1 \mu_{X_1} + ... + A_{20} \mu_{X_{20}}$ , where  $\mu_{X_i}$  was calculated using Eq. (3). The results of the reduction of  $\sigma_{\mu_{log(N_f)}}$  are summarized in Table 5.

Group  $G_2$  had the highest reduction in  $\sigma_{\mu_{\log(N)}}$  followed by Groups  $G_1$ ,  $G_4$  and  $G_3$ . If one additional experiment was added to Group  $G_2$ ,  $G_1$ ,  $G_3$ , and  $G_4$ , the reduction would be 3.75%, 0.13%, 0.08%, and 0.05%, respectively.

Next, the optimal allocation methodology was applied with 40 particles, 40 iterations. Table 6 summarizes the results after running the analysis four separate times.

Table 6. Results (Case C-1)							
Analysis	1	2	3	4			
$D_{G_1}$	0	0	0	0			
$D_{G_2}$	4	4	4	4			
$D_{G_3}$	0	0	0	0			
$D_{G_4}$	0	0	0	0			
$\sigma^{orig}_{\mu_{\log(Nf)}}$	0.036	0.036	0.036	0.036			
$\sigma^{opt}_{\mu_{log(Nt)}}$	0.031	0.031	0.031	0.031			
% Reduction	14%	14%	14%	14%			
$\Sigma C_{G_i} D_{G_i}$	\$19,240	\$19,240	\$19,240	\$19,240			

In all cases, the standard deviation of  $\mu_{\log(N_f)}$  was reduced by approximately 14% by adding 4 experiments to Group  $G_2$ . The PDF of  $\mu_{\log(N_f)}$  is shown in Figure 2. The red area with black dashed line represents  $\mu_{\log(N_f)}$  before adding any data and the blue area represents the PDF after adding the optimal experimental data.



Figure 3 and Table 7 shows the behavior of  $\sigma_{\mu_{\log(Nf)}}$  as a function of funds available. The increase in funds available implied more possible additional experiments. With more additional experiments the reduction of  $\sigma_{\mu_{\log(Nf)}}$  was higher. The decrease of  $\sigma_{\mu_{\log(Nf)}}$  as the amount of funds become available is depicted with a black dotted line, and the pink solid line shows the percent reduction of  $\sigma_{\mu_{\log(Nf)}}$  after adding experimental data to the initial data points.



Figure 3. Behavior with respect to funds available

<sup>12</sup> American Institute of Aeronautics and Astronautics

Funds	\$20,000	\$40,000	\$80,000	\$100,000	\$200,000
$D_{G_1}$	0	1	2	7	8
$D_{G_2}$	4	8	16	18	34
$D_{G_3}$	0	0	0	0	4
$D_{G_4}$	0	0	0	8	11
$\sigma^{ m \it orig}_{\mu_{\log(Nf)}}$	0.036	0.036	0.036	0.036	0.036
$\sigma^{opt}_{\mu_{\log(Nf)}}$	0.031	0.029	0.027	0.026	0.023
% Reduction	14%	19%	26%	28%	36%
$\Sigma C_{G_i} D_{G_i}$	\$19,240	\$39,326	\$78,652	\$99,854	\$199,409

Table 7. Results for different funds available

# Case C-2. Minimize $\sigma_{\sigma_{log(Nf)}}$

In this case, the objective was to minimize the standard deviation of  $log(N_f)$  standard deviation,  $\sigma_{\sigma_{log(N_f)}}$ . The standard deviation  $\sigma_{log(N_f)}$  was calculated as  $\sigma_{log(N_f)} = \sqrt{\sum_{i=1}^{20} \sum_{j=1}^{20} A_i A_j \Sigma_{ij}}$ . The analysis was conducted four separate times with 40 particles, 40 iterations. A summary of the results is shown in Table 8.

Table 8. Results (Case C-2)							
Analysis	1	2	3	4			
$D_{G_1}$	0	0	0	0			
$D_{G_2}$	4	4	4	4			
$D_{G_3}$	0	0	0	0			
$D_{G_4}$	0	0	0	0			
$\sigma^{\scriptscriptstyle orig}_{\sigma_{\scriptscriptstyle N\!f}}$	0.023	0.023	0.023	0.023			
$\sigma^{\scriptscriptstyle opt}_{\sigma_{\scriptscriptstyle N\!f}}$	0.018	0.018	0.018	0.018			
% Reduction	22%	22%	22%	22%			
$\Sigma C_{G_i} D_{G_i}$	\$19,240	\$19,240	\$19,240	\$19,240			

The reduction was approximately 22% in each of the analysis. The maximum additional experiments were allocated in Group  $G_2$  to obtain this reduction. The PDF of  $\sigma_{log(Nf)}$  is shown in Figure 4. The red area with black dashed line represents the PDF of  $\sigma_{log(N_f)}$  before adding any data and the blue area represents the PDF after adding the optimal experimental data.



A study on the sensitivity of the fretting fatigue random variables was developed by Golden et al. [5]. They showed that friction coefficient and partial slip slope, group  $G_2$ , was dominant, followed by the pad profile, group  $G_4$ . They concluded that the traditional variables in fatigue, the crack growth rate,  $G_3$ , and initial crack size,  $G_1$ , were not significant in terms of the contribution to the output variance. Instead, the friction coefficient and partial slip slope were dominant as expected in a fretting fatigue analysis. This finding supports in part the results obtained with the optimal allocation methodology, in which every analysis returned the allocation of additional experimental data to group  $G_2$ , friction coefficient and partial slip slope.

### V. Conclusions

Statistical moments obtained from simulation, i.e., mean and standard deviation of the output, often involve significant uncertainty due to the random nature of the input variables. In reliability analysis, the quantification of uncertainty is of vital importance for decision-making, where the decisions may be affected by the lack of confidence in the input variables. The optimal allocation methodology proposed here reduced the variance of the output moments. The output moments were calculated using the input population moments, which were simulated using realizations of the multivariate t-distribution and Wishart distribution.

In the optimal allocation method, the variance of the output moments may be reduced by allocating resources to obtain more experimental data of the input variables to better characterize the moments of the input probability density function. The objective of the optimization model was to minimize the standard deviation of the output moments, where the number of additional experiments was constrained to the funds available. The methodology combined a single-objective optimization algorithm with a nested-loop arrangement. The optimization algorithm used particle swarm optimization (PSO) modified to handle integer variables.

A fretting fatigue problem was explored to assess additional experiments to reduce the variance in the mean and standard deviation of cycles to failure. The number of additional experiments to add for each random variable necessary to reduce the standard deviation of the output moments depended upon several factors: the number of initial data points, the influence of the input variables, the cost of each additional experiment, and the variance of the sample mean.

In the fretting fatigue example the results found by Golden et. al [5] supported the results of the optimal allocation methodology can be used as a tool to help improve the confidence of the output moments.

### Acknowledgements

This research effort was funded in part under a grant from the National Science Foundation (HRD-0932339 through the CREST Center for Simulation, Visualization & Real Time Computing)

14

## References

- [1] Ang A and Tang W. Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering. 2nd Edition ed.: John Wiley & Sons, Inc.; 2007
- [2] Brown JM and Grandhi RV. Probabilistic Gradient Kriging to Efficiently Predict Failure Probability Confidence Intervals. 49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference; 2008; Schaumburg, Illinois; 7 - 10 April 2008. AIAA 2008-1715
- [3] Cruse TA and Brown JM. Confidence Interval Simulation for Systems of Random Variables. Journal of Engineering for Gas Turbines and Power. 2007;129(3):836-842
- [4] Ferson S and Ginzburg LR. Different methods are needed to propagate ignorance and variability. Reliability Engineering & System Safety. 1996;54(2-3):133-144
- [5] Golden, PJ, Millwater HR and Yang X. Probabilistic Fretting Fatigue Life Prediction of Ti-6Al-4V. International Journal of Fatigue. 2010;32(12):1937-1947
- [6] Kennedy J and Eberhart R. Particle Swarm Optimization. Neural Networks. 1995;4:1942-1948
- [7] McFarland JM and Riha DS. Uncertainty Quantification Methods for Helicopter Fatigue Reliability Analysis. American Helicopter Society 65th Annual Forum; 2009; Grapevine, Texas; May 27-29
- [8] Mehta SR, Cruse TA and Mahadevan S. Confidence Bounds on Structural Reliability. 34th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference; 1993; La Jolla, California; April 19-22. AIAA-93-1377-CP
- [9] Singh G and Grandhi RV. Mixed-Variable Optimization Strategy Employing Multifidelity Simulation and Surrogate Models. AIAA JOURNAL. 2010;48(1):215-223
- [10] Torng TY and Thacker BH. Confidence Bounds Assessment for Probabilistic Structural Reliability Analysis. 33rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference; 1992; Dallas, Texas; April 13-15. AIAA-92-2409-CP
- [11] Torng TY and Thacker BH. An Efficient Probabilistic Scheme for Constructing Structural Reliability Confidence Bounds. 34th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference; 1993; La Jolla, California; April 19-22. AIAA-93-1627-CP
- [12] Trelea IC. The Particle Swarm Optimization Algorithm: Convergence Analysis and Parameter Selection. Information Processing Letters. 2003;85:317-325
- [13] Urbina A, Mahadevan S and Paez TL. Resource Allocation using Quantification of Margins and Uncertainty. 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference; 2010; Orlando, Florida; April 12-15. AIAA 2010-2510
- [14] Venkataraman S, Sirimamilla RR, Mahadevan S, Nagpal V, Strack B and Pai SS. Calculating Confidence Bounds for Reliability Index to Quantify Effect of Distribution Parameter Uncertainty. 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference; 2007; Honolulu, Hawaii; 23 - 26 April 2007. AIAA 2007-1940
- [15] Walpole RE and Myers RH. Probability and Statistics for Engineers and Scientists. Fifth Edition ed. New York: Macmillan Publishing Company; 1993