Data Analysis & Statistical Methods for Command File Errors

Leila Meshkat, Bruce Waggoner, Larry Bryant

Jet Propulsion Laboratory, California Institute of Technology

This paper explains current work on modeling for managing the risk of command file errors. It is focused on analyzing actual data from a JPL spaceflight mission to build models for evaluating and predicting error rates as a function of several key variables. We constructed a rich dataset by considering the number of errors, the number of files radiated, including the number commands and blocks in each file, as well as subjective estimates of workload and operational novelty. We have assessed these data using different curve fitting and distribution fitting techniques, such as multiple regression analysis, and maximum likelihood estimation to see how much of the variability in the error rates can be explained with these. We have also used goodness of fit testing strategies and principal component analysis to further assess our data. Finally, we constructed a model of expected error rates based on the what these statistics bore out as critical drivers to the error rate. This model allows project management to evaluate the error rate against a theoretically expected rate as well as anticipate future error rates.

Background

Command File Errors (CFEs) are one of the main contributors to operational problems on space missions. Managing CFE's entails making good decisions about activities that are aimed at keeping them within acceptable bounds and preventing the occurrence of critical CFE's within reason.

The Jet Propulsion Laboratory (JPL) has been exploring the means to manage CFE's over the last fifteen years [1,2,3,4], and especially, since the loss of the Mars Climate Orbiter[5]. The management approach has been to define, collect, organize and classify the data associated with CFE's and calculate the rates for CFE's on the various types of missions, which provides a basis for comparison. The average rate for the CFE's has been assessed as 0.5%, with most of them being benign. While this metric is important to have, comparing the actual rates to this rate does not provide a full assessment of the problem as the rate can vary during the different phases of the mission and the occurrence of an error is a random variable which should be treated accordingly for correct assessment.

Our study takes a deeper look at the problem and develops appropriate stochastic models to represent the CFE's as random variables which are a function of a set of variables. We will use the following definitions in our work:

Definitions

SCMF (Spacecraft Message File) - the binary file sent to the spacecraft - common to most missions

- SCMFs can contain one to thousands of commands, usually multiple commands form sequences

Commands are the discreet instructions issued to the spacecraft by the ground; these can be real-time or sequenced

- Some commands are simple and some can be very complex due to parameterization

Sequences are sequences of commands

- Some sequences are simple and some can be complex

Blocks are groups of commands that can be reused

- Blocks can behave differently if they are parameterized
- Blocks can be simple or complex

Activity Level - A subjective estimate of the workload on the flight team. This factor was determined by averaging the number of concurrent sequence development activities from the development schedule.

Novelty - A subjective estimate of the 'newness' of flight activities being executed on the spacecraft. For our purposes, novelty should is usually defined to be a minimum of one month in duration, but usually more. For example, the first three months after launch or a mission phase transition would almost universally be a period of novel operations**.**

Operations Best Practices and the Data Set

There are certain unspoken best practices that are popular in the operations community even though they had not been carefully measured or verified. By mining the frequency of command errors, SCMF radiation, command execution, block usage, activity level and novelty levels, we are able to quantitatively analyze some of these practices. We were also able to address the question of whether these practices are true and build analytical models and tools to assess and predict CFE's in a more rigorous and objective manner. Here, we list some of those practices and in the following sections discuss the empirical analysis methods used for assessing them:

- If the flight team radiates a larger number of SCMFs, they will make more errors

    - This has always been an argument for sequencing versus real time commanding
- If the flight team issues a larger number of commands, they will make more errors
    - We are usually sensitive to not sending extraneous command to the spacecraft
- If the flight team is experiencing a high level of development activity, they will make more errors
    - This is just human factors common sense
- If the flight team uses blocks they will make fewer errors
    - Reusing blocks will reduce errors since you are using a proven product
- If the flight team is doing something novel on the spacecraft, they will make more errors
    - This is also human factors common sense

Data Analysis

The data sources available for analysis include the following:

- Information about the command files sent to the spacecraft, in terms of the number of blocks, commands and files in each month of the mission and the number of CFE's observed.
- Information about the level of activity (in terms of the products produced) by the flight team during each month.
- Information about the novelty levels associated with each month of the mission.
- General information in the databases in terms of the various errors that have occurred during the lifecycle of the flight projects and details about their causes and mitigations.

Our main goal in data analysis was to validate the results of the sigma tool. This is the first of the tools to be adopted by flight teams as it is the simplest to use. The sections below describe the different types of analysis conducted for this purpose.

Correlation Analysis

At the offset of the study, we decided to look into the correlations between the following variables in each month of the mission:

- Command File Errors
- Number of files
- Number of commands
- Number of blocks
- Activity levels during the month (in terms of products produced by the flight teams.

For the flight project in question, the highest correlation was obtained between the CFE's and the number of files sent to the spacecraft each month. The correlations with the number of blocks and number of commands was also significant, but not as high. Note that each file is

typically composed of a number of commands which are divided into blocks, so these variables are themselves highly correlated. The surprising fact was that the activity levels were not significantly correlated with the CFE. This was contrary to our expectation as we had used activity levels as a proxy for the stress level of the operators which in turn would affect their cognitive abilities. Initially we thought this might be due to the fact that the products completed in a month were not necessarily created in that month so we experimented with shifting the time for the activity levels, but it did not improve the correlation levels.

As it turns out, the activity levels, which were measured as a function of the products created during each interval, were not good indicators of the actual level of stress on each operator. This is due to the fact that the products created were often offset by the level of staffing. So projects try to allocate resources to tasks proportionately to the tasks being performed. Table 1 shows the results of our initial correlation analysis.

	Cmd Errors	SCMFs	Activity Level	Commands	Blocks
Cmd Errors	1				
SCMFs	0.571368518	1			
Activity Level	0.22474755	0.380494123	1		
Commands	0.3104797	0.23939474	0.298713416	1	
Blocks	0.296127799	0.052130687	0.268910971	0.822762979	1

Table 1: Correlation Analysis for Sample Project

Since there really is a distinction between the different phases of the mission, we experimented with partitioning the data based on the phases of the mission and conducting both the correlation and regression analysis on the various phases of the mission. But some phases are very short and sufficient data does not exist for them, so the results of the experiments were non-conclusive.

Later, we added the "Novelty Level" as another variable and realized that the correlation between CFE's and Novelty levels were the second highest after the correlation between CFE's and the number of files. Table 2 includes the results of the correlation analysis after adding the novelty factor.

	Cmd Errors	SCMFs	Activity Level	Commands	Blocks	Novelty Level
Cmd Errors	1					
SCMFs	0.57136852	1				
Activity_Level	0.22474755	0.380494	1			
Commands	0.3104797	0.239395	0.298713	1		
Blocks	0.2961278	0.052131	0.268911	0.82276298	1	
Novelty	0.53297418	0.579055	0.44023	0.36065182	0.345509	1

Table 2: Correlation Analysis after adding the Novelty Factor

Regression Analysis

The next data analysis technique used was a regression analysis. The goal of this analysis was to determine how much of the variability in the Command File Errors can be explained with a nonlinear function of the variables in question. Of course the caveats of this study are that (1) the CFE's are not continuous variables and therefore they can't be predicted as a continuous function of the variables and (2) we know that the behavior of the system is probabilistic rather than quadratic, so a quadratic equation does not completely capture the variability of the CFE's.

For the flight projects analyzed, the R-squared value was approximately 50% based on all the variables. When we conducted a regression analysis only on the Novelty factors and the number of files, the R-squared value was reduced to about 40%. Given the caveats listed above, we conclude that the variables in question, especially the novelty levels and the number of files are highly significant factors.

There are still two other issues to assess: one is the suitability of the Binomial Distribution for representing the behavior of the errors and second is the variables which were deemed significant (number of commands and number of blocks) but are highly correlated with the number of files (the variable deemed to be the most significant.) We use a chi-squared test to assess the appropriateness of the Binomial Distribution and Principal Component Analysis to find an "Adjusted number of files" variable which takes into consideration the contribution of the blocks and commands as well.

Chi-squared Test

We used a chi-squared goodness of fit test [7] to test the hypothesis that a Binomial distribution with the parameters we had had identified was in fact the correct distribution for the distribution of the command file errors. For this purpose, we binned the files into those with errors and those without errors. We then calculated the expected number of files with errors based on our distribution, and also assessed the observed number of errors in each case. Then, using the equation$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$, where Oi is the observed frequency for bin i, and Ei is the expected frequency for that bin based on the hypothesis that our distribution is correct, we obtain the value to be compared with a chi-squared distribution with one degree of freedom to assess the goodness of fit.

The reason for considering only one degree of freedom is that we have two bins (error or no errors) and the total number of files are constrained, so we have only one degree of freedom. Consulting with the Chi-squared distribution, we see that the probability of getting the number of errors that we did, for high novelty operations or the p-value is 0.47. This is significantly higher than the minimum criteria for statistical significance (which is typically between 0.001 and 0.05) and therefore we cannot reject the null hypothesis that our Binomial distribution is

the correct distribution. One interesting observation after conducting the Chi-squared test was that even though a Binomial distribution with a rate of 0.004 for low-novelty operations had a reasonable p-value, the p-value improved significantly when we set the rate to 0.005. So this test can be used as a means for calibrating the distribution as well.

Principal Component Analysis

The inputs to the Binomial Distribution are the number of (files, commands, blocks) and the error rate per (file command, or block). So we can only work with two variables. The novelty factor is taken into consideration by considering different error rates for different novelty levels. But for the total number of (files, commands, blocks) radiated to the spacecraft, we need a single variable that contains the information embedded in those three variables.

Correlation analysis

	Files	Commands	Blocks
Files	1.00000000	0.2393947	0.05213069
Commands	0.23939474	1.0000000	0.82276298
Blocks	0.05213069	0.8227630	1.00000000

Importance of components:

	Comp.1	Comp.2	Comp.3
Standard deviation	1.368238	0.9859539	0.39474034
Proportion of Variance	0.624025	0.3240351	0.05193998
Cumulative Proportion	0.624025	0.9480600	1.00000000

Loadings:

	Comp.1	Comp.2	Comp.3
Files	0.233	0.959	0.160
Commands	0.700	-0.712	
Blocks	0.675	-0.277	0.684

Figure 1: Output from the R Principal Component Analysis

Since these variables are highly correlated, we use a Principal Component Analysis [6], using the statistical analysis tool, R [8], to determine the key linearly uncorrelated variables and the contribution of each of these variables for capturing the variability in the data. Once this assessment has been conducted, we can find the original variables as linear functions of the principal components. Therefore using this reverse transformation, we can determine a linear combination of the original variables which captures most of the variability of the original data. This combination guides is in finding an "Adjusted number of files" for our Binomial Distribution. The results of the analysis conducted in R are shown in figure 1. As we see, the first principal component explains 0.62 of the variability and the second principal component 0.324 of the variability. The contribution of the third principal component is not significant (0.05). In the loadings section of the table, we find the principal components as functions of the original variables.

So Comp 1=0.233xFiles +0.7xCommands+0.675 x Blocks and

Comp 2=0.959 x Files -0.712 x Commands -0.277 x Blocks.

If we consider the amount of variability that each of the principal components describe as the weight of that component, then we can find the weighted sum of the principal components to be:

0.624 x Comp 1 +0.324 x Comp 2

= 0.624 (0.233xFiles +0.7xCommands+0.675 x Blocks) + 0.324 (0.959 x Files -0.712 x Commands -0.277 x Blocks)

= 0.456 x Files +0.206 x Commands +0.33 x Blocks

This is an optimal equation for explaining the variability in the three variables of files, commands and blocks. Note however, that the correlation between these variables and the CFE's is at most close to 50%. Therefore it does not necessarily optimize in terms of finding the "Adjusted number of files". Interestingly, when we use this equation to find the number of trials for the Binomial distribution and conduct a chi-squared test to determine whether or not it is an appropriate distribution, the null hypothesis is rejected. In other words, the distribution does not accurately represent our CFE rates.

Nonetheless, this equation gives us an idea about the relative contribution of each of these variables. Starting with this equation as an initial function and refining it by trial and error, we obtain the following equation for the "adjusted number of files"

$$0.9x \text{ Files} + 0.005 \times \text{Commands} + 0.095 \times \text{Blocks}$$

This equation achieves a higher correlation with CFE rates as compared with the actual number of files. It also produces higher p-values in the Chi-squared goodness of fit analysis as compared to the actual number of files.

The rationale for this is that even though the number of files are the most significant indicator amongst the three variables, all files are not equal. Files that have more commands and blocks in them are more likely to have errors.

From a practical standpoint, it requires a significant effort to mine the number of blocks and the number of commands from files that are sent to the spacecraft. Therefore, it may be reasonable to simply use the number of files at the initially. If it appears that the CFE rates are outside of the acceptable ranges, then one area to analyze further is the characteristics of the files being transmitted, in terms of their associated block and commands.

Sigma Tool

We have used EXCEL to model a Binomial distribution which describes the expected CFE rate as a function of the number of files that are sent to the spacecraft and the novelty level. Using this model we can assess whether or not the deviations that are observed in these rates are within scope or if these deviations indicate some type of discrepancy that needs to be addressed.

Because the model provides the standard deviations from the expected rate, we call it 'sigma-tool'. Figure 2 shows the resulting graph for the sigma tool for our flight project. This graph helps us get an understanding of the error rates on a month by month basis and is used by management for the purpose of assessing whether or not the observed rates require mitigating action. Using this model, we can also develop expectations for the error rates of the flight project in future months in terms of the level of novelty of the mission and the number of files that are expected to be sent to the spacecraft.

Binning the data in various ways (such as looking at quarterly error rates) also provides additional perspectives for analyzing the data. This can be useful because error rates are often very low.

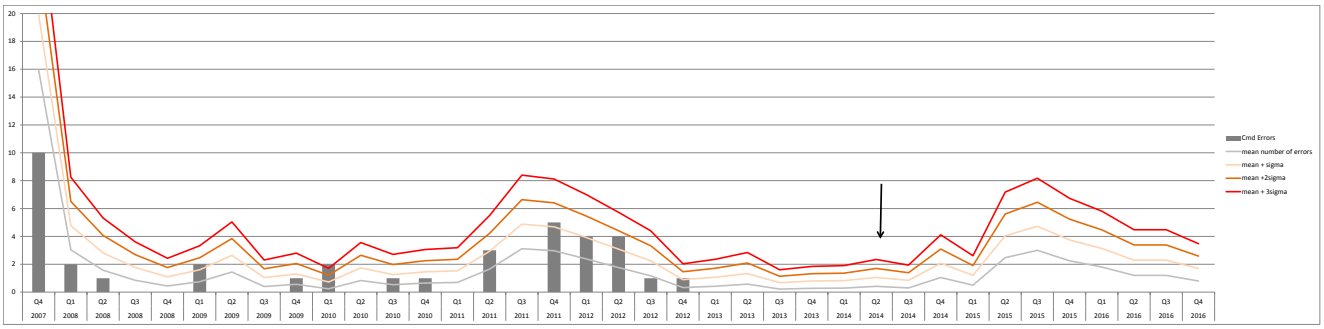

Figure 2: Sigma Tool for a Sample Mission.

The input to the Sigma Tool is the number of files sent each month and the expected level of novelty for the operations during that month. We also consider tailoring the error rates used by the Binomial distribution to the particular mission based on the experience of the Mission Operations Assurance Manager (MOAM) and the range of error rates from previous missions. The MOAM can tweak the rate based on their assessment of where the mission in question falls in the context of similar past missions. For example, if a project has had a long term error rate that is higher than a typical JPL mission, the model use that rate. In this case performance against past mission history is displayed. However if we used the .5% rate mentioned earlier across many missions, the projects performance against this mean would be displayed by the model. Both may be valuable, but care needs to be taken to make sure users understand what the

Summary & Conclusions

In this paper, we briefly described the background work related to using probabilistic modeling techniques, in particular Bayesian Belief Networks, to manage Command File Errors. We then discussed the distribution of command file errors as a function of several key variables and used a variety of empirical analysis techniques to refine and validate these distributions. The distribution used for this purpose is a Binomial Distribution. The parameters for this distribution are obtained by considering the number of trials to be a function of the number of files and the number commands and blocks in each file. The error rate is dependent on the level of novelty of the operations and is derived from similar data from past missions by the Mission Assurance Operations Manager. Building this model has resulted in a tool, which we call the “Sigma Tool”. This tool is used by management to assess the range of the CFE’s and determine what to expect based on mission plans. If the expected error rates seem unacceptable, then our Bayesian Belief Network (BBN) model, which we have previously reported on, is used to conduct analyses to determine methods for decreasing them. This tool can also help determine whether or not the observed error rates are within the scope of our expectations. Using this tool on one particular mission led us to realize that something was off

and in further perusal of the mission it became clear that the data collection by our spacecraft contractor was incomplete – hence the anomalous behavior in our results.

The next steps are to further infuse the tool developed based on the distribution into flight project teams to gain a wider range of understand of how the model performs.

Finally, here is our assessment of the best practices listed in the background section:

- If the flight team radiates a larger number of SCMFs, they will make more errors
    - True! This seems to be the primary driver for CFEs
- If the flight team issues a larger number of commands, they will make more errors
    - True, but not as important as SCMFs radiated or novelty
- If the flight team is experiencing a high level of development activity, they will make more errors
    - False, it appears staffing takes this risk down, in other words we manage this one on Dawn
- If the flight team uses blocks they will make fewer errors
    - Not true. We suspect that blocks are a double edged sword as they are sometimes used 'inappropriately' on Dawn (screwdrivers don't make good hammers)
- If the flight team is doing something novel on the spacecraft, they will make more errors
    - True! The second biggest driver.

Acknowledgements

The work reported in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. The authors would like to acknowledge the material provided by Grant Faris as well as his invaluable insights and advice.

References

1. Grant Faris, "Proposed Institutional Command File Error Definition, Categories, and Metrics". 3 April 2008.
2. G. B. Faris, L. W. Bryant "Improving Operations: Metrics to Results".
3. Kelly Bezjak and Bruce Waggoner, "Mission Operations Flight School", 2010.
4. H. Kwong-Fu, C. Scott, R. Smith, G. Chin, " Command File Error Metrics Working Group Analysis & Results", 2008.
5. L. Meshkat, L. Bryant, B. Wagoner & R. Thomas, "Soft Factors & Space Mission Failures: Quantifying the Effect of Management Decisions", AIAA Science & Technology Forum, January 2014.

6. Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4
7. Plackett, R. L. (1983). "Karl Pearson and the Chi-Squared Test". International Statistical Review (International Statistical Institute (ISI)) 51 (1): 59–72
8. The R Project for Statistical Computing: http://www.r-project.org/