An Early Look at COQUALMO in the JPL Environment

John D. Powell
John.Powell@jpl.nasa.gov

California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena CA 91109-8099

Abstract

This paper will describe the initial efforts to validate and calibrate the COQALMO model within the JPL environment. The risk to JPL software projects precipitated by higher than expected defect densities late in the software lifecycle impact spacecraft/missions in several ways. Among them are risks of missing "hard" schedule deadlines, which may be dictated by immovable launch dates and the timing of critical spacecraft encounters with objects of interest in space. Further risks include, but are not limited to, failures during execution of mission critical software that has only one opportunity to perform its functionality correctly. For example, software critical to mission critical non-repeatable phases such as Entry Decent Landing (EDL) software and instrument and data collection software during mission encounters. The ability to predict expected defects densities early in the life cycle with models such as COQUALMO will allow project managers to plan schedules and resources more efficiently to ensure the delivery of high quality software within the critical time constraints discussed above. However, as with any predictive statistical model like COQUALMO, it must be calibrated to the environment in which it is to be used (i.e. JPL) to provide maximum benefits to software projects that come to rely on the information it provides.

1. Introduction

The COQUALMO model, an extension to the COCOMO II model, seeks to predict software defect densities based on early lifecycle characteristics. [1,2,3,4] The characteristics include the same input data for the COCOMO II model, namely 21 of the 22 cost drivers. In addition to the COCOMO II cost drivers, three defect removal profiles must be collected from software projects. The defect removal profiles describe the defect removal activities to be performed during a given software project and the degree of rigor with which they will be applied. The COQUALMO model then produces a prediction of the introduction and removal of requirements, design and code & test defects that will occur as well as the residual delivered defect density. The current COQUALMO model is based on expert opinion though two rounds of Delphi analysis conducted by the Center for Software Engineering (CSE) at the University of Southern California (USC). [5,6]

2. Methodology

This paper discusses the experimental effort that is underway at the Jet Propulsion Laboratory (JPL) to tailor this model for maximum effectiveness in predicting defect rates for flight and ground software for spacecraft. The fist step of this research has been to examine past projects through: 1) The collection of COCOMO II cost drivers and defect removal profiles through interviews with project managers and other personnel still available at JPL. 2)

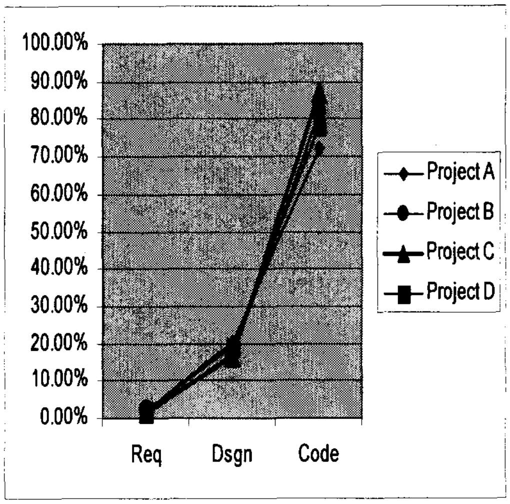

Figure 1: Predicted of Defect Introduction by Category

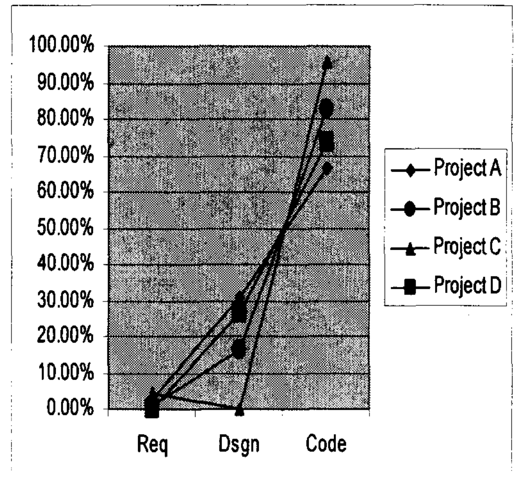

Figure 2: Actual Defect Introduction by Category

Examination and proper categorization of actual defect data from the past projects for which data from step 1 is available. This step will yield defect data from JPL projects in a form that reasonably is comparable to the COQUALMO model's defect prediction output. The second step is to validate the COQUALMO model within the JPL environment using these results. This is being done by analysis of the categorized defect data in conjunction with the corresponding COQUALMO predictions to determine whether or not consistent patterns of correlation and or deviation exist.

Currently, these fist two steps are being performed in parallel as defect data and COQUALMO input parameters become available for past projects. As of the writing of this paper 4 JPL-COQUALMO data points exist. Additional raw data exists which may be used to construct partial data sets for COQUALMO and expanded to full data points in the near future. However, this paper will exclusively discuss the 4 data points constructed from complete data sets.

A valid COQUALMO data point consists of sufficiently complete and verifiable:

- COQUALMO prediction outputs based on model input parameters that are verified through competent personnel from the past project
- Actual defect data that is based on documented defect-tracking data from the past project's development archive. When necessary relevant past project personnel are contacted to clarify context and interpretation of this data.

3. Results

The 4 COQUALMO data points in existence at JPL indicate commonalities in the correlation and deviation from prediction results produced by the CSE's current COQUALMO implementation. The nature of these commonalities will be discussed in the following two subsections (3.1 and 3.2). This discussion is restricted to the context of the ongoing COQUALMO validation work in relation to the JPL environment with the caveat that 4 data points is insufficient to statistically valid statements about trends of correlation.

3.1. Current Correlations

The projects represented by the 4 data points in existence represent a wide cross section of the types of software projects developed at JPL. Both Flight and ground software is represented as well as software designed for vastly different missions within these categories.

Figure 1 illustrates that the predicted percentage of defect introduction rates

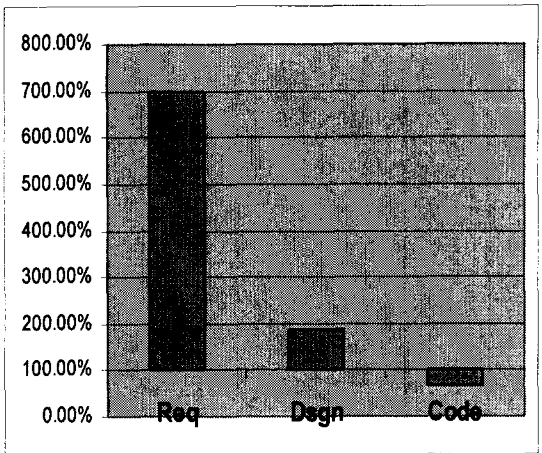

Figure 3: Defect Prediction Error by Category

generated by COQUALMO across defect categories is similar for all four projects. This indicates that COQUALMO does not recognize a distinction between the variations in the types of software project sampled thus far at JPL. However, when the actual distribution defects across the defect categories for are examined (Figure 2), it is apparent that there is in fact little deviation across the projects with respect to this metric. Thus, COQUALMO's correlation in prediction is largely justified in this respect. It is important to note that there is an outlying feature in the actual defect data for Project C because no (zero) design defects are reported. This is regarded as highly unusual and is potentially attributable to errors in root cause determination of defects for the purpose of categorization. This phenomenon is still being investigated to the extent possible for this past project.

3.2. Current Deviations

COQUALMO's discrimination between defects of differing types (requirements, design, code and test) is surprisingly accurate when compared with actual defect rates from this four-project JPL sample considering no local calibration of the model was attempted before obtaining these results.

COQUALMO dramatically over estimates the total number of defects in relation to the number of defects discovered in the four JPL projects that have been sampled thus far. Figure 3 illustrates COQUALMO's rate of defect prediction error within each defect category. The average number of design defects (~ +188%) and code & test (~ - 65%) defects estimated by COQUALMO is within reasonable constraints for an uncalibrated model. However, the predicted number of requirement defects (~ + 700%) is greatly overstated. The degree of this over statement in light of the relatively accurate discrimination between defect types within the overall number of defects predicted by COQUALMO (See Figure 1) is due to the fact that the number of actual requirements defects was proportionally small. Thus the 7X over statement of requirements defects represents a much small number of falsely predicted requirements defects in absolute terms than in the case of design and code & test defects.

4. Implications on Future Validation

Under the hypothesis that these commonalities and consistencies may continue to persist over a larger set of JPL's COQUALMO data points, a discussion of their implications with respect to future calibration of COQUALMO to the JPL environment is offered.

Various calibration strategies emerge under the conditions arising out of the current sample data (given the hypothesis above). First and foremost is the need to realign the to total number of predicted defects to more closely follow the total number of actual defects in the software. However, this must be done in a manner that largely preserves the allocation percentages among the defect categories because that aspect of the COQUALMO prediction is relatively accurate. Conversely, investigation into the volume of significant defects that are handled informally (if any) and not recorded into defect track for metrics purposes must be considered.

The notions surrounding the phenomenon where by a series of code & test defects are recorded before finding that, in reality, the set represents a design defect that remained undiscovered during his time period must be addressed. One approach may be to reclassify the code defects as a design defect and use the new data as part of the baseline. However, a more complete picture, which will provide better information needed by JPL software managers for planning purposes, includes the integration of rates at which design defects continue on to the coding phases and the amount of effort expended in the process of addressing resulting code & test defects before discovery of the true root cause. By knowing that this situation occurs at a given rate and costs projects some corresponding amount of time and money, resources may be planned in advance to:

- Deal with this eventuality to mitigate surprises late in the project.
- Justify the allocation of additional resources and employment of new techniques during the design phases to avoid the more expensive alternative

5, Future Work

Future work will consist of: 1) the collection of data from ongoing projects to produce additional COQUALMO data points for future calibration purposes, 2) refinement of the COQUALMO model as needed to provide maximum benefit to JPL projects, 3) refinement of the data collection process and effective integration of this function into the internal JPL processes in order to minimize intrusion on JPL software projects in the future, 4) working with JPL software projects to define and provide effective services and analysis results that will benefit ongoing JPL software projects throughout the development lifecycle. These services include determination of relevant correlations between software defects and cost and schedule factors for JPL projects as well as reliable predictors of risk in these areas. Preliminary plans in these areas of future work have been underway at JPL and are currently being deployed through the Software Quality Improvement (SQI) project.

It bears noting that the large overlap between COCOMO II and COQUALMO input parameters combined with close collaboration with the software cost estimation functions at JPL has already allowed the COQUALMO validation and calibration effort to provide an immediate value added to ongoing JPL software projects in the form of software cost estimation services. The ability to offer an immediate benefit along with the experimental infusion of technology has been invaluable in securing project cooperation commitments for the COQUALMO research.

6. Acknowledgement

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] Barry Boehm et al., Software Cost Estimation with COCOMO II, Prentice Hall Inc., Upper Saddle River, NJ 07458, 2000

[2] Sunita Devnani-Chulani, Bayesian Analysis of Software Cost and Quality Models, Ph.D. Dissertation, University of Southern California 1999

[3] Sunita Devnani-Chulani and Barry Boehm, Modeling Software Defect Introduction Removal: COQUALMO (Constructive QUALity MOdel), USC-CSE-99-510

[4] Sunita Devnani-Chulani, Modeling Software Defect Introduction Proc. California Software Symposium '97, USC-CSE-98-503

[5] Sunita Devnani-Chulani, Incorporating Bayesian Analysis to Improve the Accuracy of COCOMO II and Its Quality Model Extension, USC-CSE-98-506

[6] Sunita Devnani-Chulani, Results of Delphi for the Defect Introduction Model (Sub-Model of the Cost/Quality Model Extension to COCOMO II) USC-CSE-97-505