International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:4, 2015

Establishing of Function Point Process Based On Stochastic Distribution

Do Syung Ryong, Kang Hyun Su

***Abstract*—**This study aims to establish function point process based on stochastic distribution. In order to demonstrate effectiveness of the study we present a case study that it applies suggested method on an automotive electrical and electronics system software development based on Monte Carlo Simulation. It is expected that the result of this paper is used as guidance for establishing function point process in organizations and tools for helping project managers make decisions correctly.

***Keywords*—**Function Point, Monte Carlo Simulation, Software Estimation, Stochastic Distribution.

I. Introduction

THERE exist invisibility and intangibility in software development process; for instances, size, schedule, effort, cost, scope and quality. For that reason, it is possible to check the quality of the project only at the end of the project life cycle. Due to this characteristic, project managers can make decision wrong. Thus, software project activities influence these six variables as distributions rather than deterministically [1].

For successful software projects, managers systematically have to manage and balance six variables based on historical probability. Especially, software size and cost estimation activity requires enough historical data because incorrect estimation may lead project to fail.

Though software size estimation has uncertainty in initial development it has traditionally performed using deterministic method: LOC (Lines Of Code), COCOMO (COnsructive COst MOdel), FP (Function Point), SLIM (Software LIfecycle Management). However, these methods may fail to consider numbers of variable factors in the software development process [1]. Therefore, it is necessary to introduction a stochastic method to increase the accuracy of the estimation considering process variations [1].

Review of literature reveals widespread use of stochastic methods like Monte Carlo simulation, Sensitivity Analysis, Bayesian Belief Networks such as in construction [2] and financial projects [3] etc. However, use of Monte Carlo simulation in software projects especially for estimation is relatively much less [4].

Sung Ryong. Do is with Department of Quality Management, Hyundai Autron, Bundang-Gu, Seongnam-si, Gyeonggi-do, 463-400, Korea (corresponding author; phone: 82-10-2864-1897; fax: 82-31-627-0991; e-mail: SungRyong.Do@hyundai-autron.com).

Hyun Su. Kang is with Department of Quality Management, Hyundai Autron, Bundang-Gu, Seongnam-si, Gyeonggi-do, 463-400, Korea (phone: 82-10-8798-5664; fax: 82-31-627-0991; e-mail: HyunSu.Kang@hyundai-autron.com).

We proposed a function point process based on stochastic distribution and a case study based on Monte Carlo Simulation applying on an automotive E/E (Electrical and Electronic system software development. The difference with existing function point process is utilizing stochastic distribution of existing historical data as counting data functions, transaction functions, and value adjustment factors. We present a case applied to E/E control system software vendors to verify the findings.

For comparison with existing function point process, we performed the statistical analysis. In paired T-Test, the scheme proposed is more excellent than existing methods by verify accuracy of the estimate and it is proven high through the correlation analysis with LOC.

It is expected that the result of this paper is used as guidance for establishing of function point process in organizations and tool for helping project manager make decisions correctly.

The remainder of the study is organized as follows: Section II gives background theories and related works. Section III introduces a function point process based on stochastic distribution. Section IV gives a research result of applying on an automotive electrical and electronics system software development. We describe conclusion and future works in Section V.

II. Background Theories and Related Works

A. Quantitative Management

Quantitative management is to managing a project or work group using statistical and other quantitative techniques to build an understanding of the performance or predicted performance of processes in comparison to the project's or work group's quality and process performance objectives, and identifying corrective action that may need to be taken [5]. Statistical techniques used in quantitative management include analysis, creation, or use of process performance model; analysis, creation, or use of process performance baselines; use of control chart; analysis of variance, regression analysis; and use of confidence intervals or prediction intervals, sensitivity analysis, simulations, and tests of hypotheses.

Quantitative management is classified into deterministic and stochastic method, such as Fig. 1. The deterministic method is used to determine a representative value such as an average of the data, mode, and maximum value. The deterministic method is easy to use and to understand, but it is not suitable in the field requiring accurate analysis as it does not take diversity and variability of variables into account. On the other hand, stochastic approach is more focused on uncertainty of variables because it depends on the stochastic distribution extracted from historical data [1]. One of these typical methods for stochastic approach is Monte Carlo simulation technique.

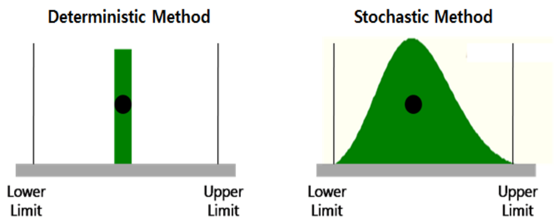


Fig. 1 Deterministic Method and Stochastic Method

Monte Carlo simulation is a technique of selecting arbitrarily from the probability distribution of values for use in the simulation, as shown in Fig. 2, also referred to as a simulated sampling technique. The advantage of the technique is to generate a random number in any condition for the input given the number of all cases and providing distribution and statistics generated from the result to support the decision-making [6].

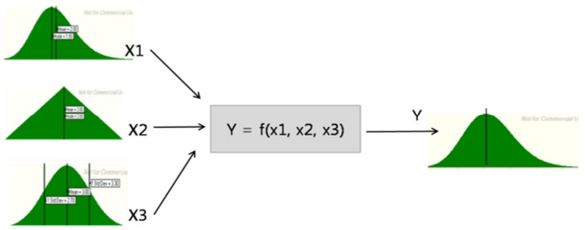


Fig. 2 Monte Carlo Simulation Concept

Typical studies of the use of Monte Carlo simulations on the software development are as follows: Reference [7] presented a case applying a Monte Carlo simulation technique for sensitivity analysis of the COCOMO II model. Reference [6] proposed a Monte Carlo simulation-based risk management process and presented the simulation results to the project schedule prediction. Reference [4] developed a probabilistic model for process variations notation for software development step by effort prediction and proposed a method for verifying utilizing Monte Carlo simulation.

Although it recently presented a variety of statistical techniques and guidelines, including the Monte Carlo simulation and quantitative management for the SEI (Software Engineering Institute) in the United States, the software sector is still an early stage and, in particular, estimation category needs further study.

B. Estimation and Function Point

Software development size and effort estimation activity has been issued in software engineering. LOC, COCOMO, FP, SLIM, and many other techniques are being studied and their research that improves the technique has currently been being conducted.

FPA (Function Point Analysis) denotes a family of algorithmic methods for size estimation. This method separately evaluates two classes of the attributes of a software system: size factors and influence factors. The first version of FPA, invented by Albrecht at IBM in 1979, proposed a new metric (i.e., function point) for software size rather than lines of code. The International Function Point User Group (IFPUG) adopted a revised method, defining a function point as a means to measure software size by quantifying the functionality provided to the user based solely on logical designs and functionality specifications. Because the functionality of a software system, from the user's perspective, usually emerges early in a project, FPA offers the unique advantage of being applicable during the early stage, when other approaches to size measurement are not appropriate.

FPA model classifies the functions of a software system into five types: internal logical files (ILF), which are internally maintained, logical groups of data; external interface files (EIF) that are passed or shared among applications; external inputs (EI), which refer to the unique user data or control inputs that add to or change the data; external outputs (EO), which are the unique user data or control outputs that fall outside the boundaries of the system; external inquiries (EQ), which are the unique input that generates immediate output. Furthermore, IFPUG groups these functions into either data functions (ILFs and EIFs) or transaction functions (EIs, EOs, and EQs).

Existing research related to function point are as follows: Reference [8] proposed a function point method of applying a regression model for the adjustment factor. Reference [9] proposed a scheme for optimizing the 14 general system characteristics based on the existing data. Reference [10] suggested estimation method of the software size based on the function point and proposed a method for predicting the productivity of specific development stage based on the results.

Several studies [8]-[10] have related function score in progress, but previous studies have proposed a method of predicting the software size using a representative value. That is, there are conducted on a deterministic way, there is a problem that does not consider the variability of the estimation factors. Therefore, it is necessary to consider the variability in the study based on the simultaneous variation of the various parameters that affect the result can be considered stochastic manner.

III. Function Point Process Based On Stochastic Distribution

This study deals with upgraded function point process by applying stochastic theory to the existing function point estimation. The key point of the process is to apply stochastic distribution theory to the main estimation elements such as data functions, transaction functions and value adjustment factors. It is suggested that statistics of main estimation element should be extracted from the same type of domain.

A. Construct Stochastic Distribution of Data Functions

1) Extract data function on the same type of domain from the result of existing function point.
2) Classify data function in ILF and EIF.
3) Construct stochastic distribution based on the result of classification.

B. Construct Stochastic Distribution of Transaction Functions

1) Extract transaction function on the same type of domain from the result of existing function point.
2) Classify transaction function in EI, EO and EQ.
3) Construct stochastic distribution based on the result of classification.

C. Construct Stochastic Distribution of Value Adjustment Factors

1) Measure general system characteristics influencing performance of software, such as data communications, distributed data processing, complex processing, reusability and facilitate change.
2) Construct stochastic distribution based on the result of measurement.

The steps for the function point process based on stochastic distribution are as Fig. 3. What is difference from existing function point process is that probability distribution theory is applied to the decision on data functions, transaction functions, and value adjustment factors.

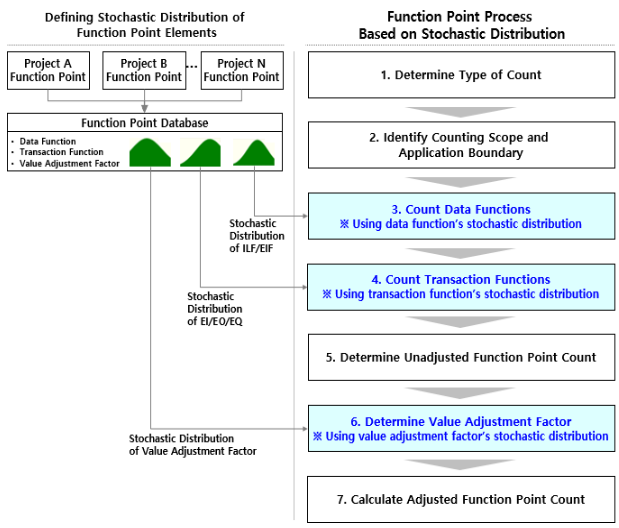


Fig. 3 Function Point Process based on Stochastic Distribution

IV. Research Result

A. Case Study Overview

1) Feature of Automotive E/E Software Development

In order to verify effect of the research we applied function point process based on stochastic distribution to domestic automotive E/E control system software development

The feature of automotive software development is that most projects are developed as enhancing functions on software already implemented when the new automotive and system are developed. For example, when ECU (Engine Control Unit) software embedded on light-weight vehicle is developed most of the functions (fuel, air, torque) on middle size vehicle are commonly reused and some parts of the functions are modified. Due to high-reusability of the functions historical data extracted from past similar projects and stochastic distribution are significant on the function estimation.

2) Introduction of Projects for Case Study

We have analyzed historical data from 17 EMS (Engine Management System) projects to define stochastic distribution for the function point estimation elements. Target projects to be analyzed are EMS for light-weight vehicle and highly reusable (80% reusability). EMS is the system to control amount of air intake, fuel and ignition timing so as to get torque requested. System structure of the EMS is as Fig. 4 and is separated into several functions such as torque, air, fuel, monitoring and so on.

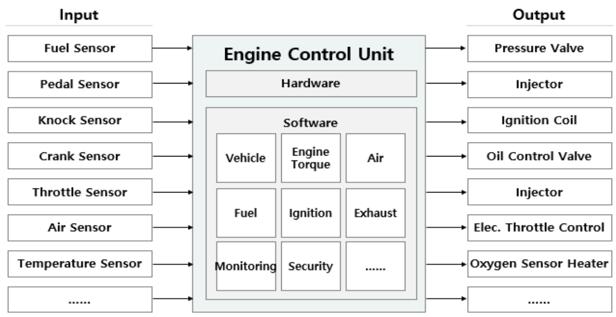


Fig. 4 System Structure of ECU

B. Case Study Result

In this case study, Monte Carlo simulation was used to apply function point process based on stochastic distribution and Crystal Ball based on Microsoft Excel is used as supporting tool.

1) Determine Type of Count

A project is the type of new development project that EMS software is attached to light-weight vehicle.

2) Identify Counting Scope and Application Boundary of ECU

Counting scope of A project includes EMS software, driver and external ECU software as Fig. 5. Application boundary is limited to the ECU.

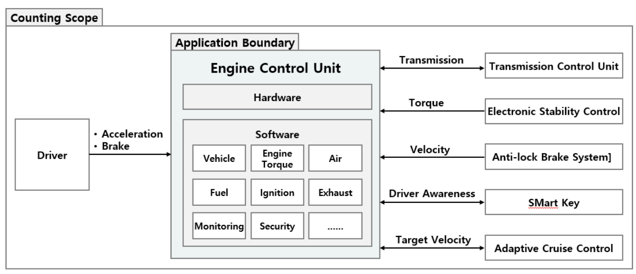


Fig. 5 Counting Scope and Application Boundary of ECU

3) Count Data Functions

Data functions of A project is measured using stochastic distribution in Table I defined by existing EMS software data functions. Respective measured data is summed according to formula, as in (1):

$$\text{Sum of Data Function} = \sum (ILFs + EIFs) \quad (1)$$

Equation (1) is defined as a formula in Crystal Ball and simulation is executed. As a result, it showed that data function of engine torque component is from minimum 60.44 to maximum 72.33 in 90% chances and average is 66.48 as described in Fig. 6 (a). In terms of sensibility, as described in Fig. 6 (d), influence of ILF is higher than influence of EIF and data function 'Torque' and 'Speed' has most powerful influence.

4) Count Transaction Functions

Transaction functions of A project is measured using stochastic distribution in Table II defined by existing EMS software transaction functions. Respective measured data is summed according to formula, as in (2):

$$\text{Sum of Transaction Function} = \sum (EIs + EOs + EQs) \quad (2)$$

Equation (2) is defined as a formula in Crystal Ball and simulation is executed. As a result, it showed that transaction functions of engine torque component is from minimum 21.89 to maximum 26.05 in 90% chances and average is 24 as described in Fig. 6 (b). In terms of sensibility, as described in Fig. 6 (e), there are not significant differences among EI, EO and EQ. Torque distribution and Idle control transaction function has most powerful influence.

5) Determine Unadjusted Function Point

Unadjusted function point of A project is measured with sum of data functions and transaction functions, as in (3):

$$\text{UFP} = \sum (ILFs + EIFs) + \sum (EIs + EOs + EQs) \quad (3)$$

Equation (3) is defined as a formula in Crystal Ball and then simulation is executed. As a result of execution, it showed that unadjusted function point is distributed from 84.10 to 96.70 in 90% chances and average is 90.47 as described in Fig. 6 (c). In terms of sensibility, as described in Fig. 6 (f), influence of data functions is higher than influence of transaction functions and influence of data function 'Torque' and 'Speed' is most powerful.

6) Determine Value Adjustment Factors

Determine value adjustment factors means general system characteristics is somewhat subjective. In order to minimize the problem specialist has decided value adjustment factors through brain storming. However, as it is still subjective value adjustment factors of A project is measured using stochastic distribution in Table III defined from existing adjustment factor of EMS software. Respective measured adjustment factor is summed according to formula, as in (4). Equation (4) is defined as a formula in Crystal Ball and then simulation is executed. As a result of execution, it showed that value adjustment factor is distributed from 1.13 to 1.18 in 90% chances and average point is 1.16 as described in Fig. 7 (a). In terms of sensibility, as described in Fig 7 (c), influence of data communications and distributed data processing is most powerful.

$$\text{Sum of VAF} = (\sum (GSCs) X\, 0.01) + 0.65 \quad (4)$$

7) Calculate Adjusted Function Point Count

The way how to measure adjustment function point of A project is to multiply unadjusted function point and adjustment factor. Equation (5) is defined as a formula in Crystal Ball and then simulation is executed. As a result of execution, it showed that adjustment function point for engine torque component is distributed from 96.67 to 112.11 in 90% chances and average point is 104.66 as described in Fig. 7 (b). In terms of sensibility, as described in Fig. 7 (d), influence of data function is higher than influence of transaction function as well as adjustment factor, and influence of data function 'Torque' and 'Speed' is most powerful.

$$\text{AFP} = UFPXVAF \quad (5)$$

C. Verification of Case Study

In order to verify effectiveness of the case study we conducted statistical analysis in phase of difference and correlation comparison.

1) Verification of Difference with Final Function Point

In general, there is a difference between function point measured at the beginning and function point at the end of the project. As described in Table IV, through the paired t-test, we could verify the difference among function point measured by traditional method, function point by the way this study proposed and function point at the end of the project. After we measure statistical significance of function point determined at the end of the project and function point measured by traditional method, there is a difference that t is t-value is 3.366 and significance probability is 0.008 in the significance level of .05 as described in Table V. On the contrary, as described in Table V, there is no significant difference that t-value is -2.150 and significance probability is .069 in the significance level of .05 as we compared function point determined at the end of the project and function point measured by using stochastic distribution. Consequently, the way this study suggested is more accurate in predicting final function point than traditional way.

TABLE I
FUNCTION POINT AND STOCHASTIC DISTRIBUTION OF DATA FUNCTIONS

Name	Type	Min.	Mean	Max.	Std. dev.	Distribution
Torque	ILF	7	12.29	15	3.10	(Normal)
Surge Damper	ILF	7	8.06	10	1.48	(Normal)
Speed	ILF	7	12.76	15	2.84	(Normal)
Coordinator	ILF	10	13.53	15	2.35	(Normal)
Air	EIF	5	7.12	10	1.87	(Normal)
Fuel	EIF	5	5.47	7	0.87	(Normal)
Ignition	EIF	7	9.12	10	1.41	(Normal)

TABLE II
FUNCTION POINT AND STOCHASTIC DISTRIBUTION OF TRANSACTION FUNCTIONS

Name	Type	Min.	Mean	Max.	Std. dev.	Distribution
Torque distribution to the path: air, fuel, ignition	EO	4	6.06	7	1.20	(Normal)
Dampening of load alternation via limitation of torque gradient	EO	4	4.24	5	0.44	(Normal)
Coordination of vehicle and engine torque demands	EQ	4	5.65	6	0.79	(Normal)
Torque limit (max. speed)	EI	3	3.29	6	0.77	(Normal)
Idle control	EI	3	5.18	6	1.01	(Normal)

TABLE III
DEGREE OF INFLUENCE AND DISTRIBUTION OF VALUE ADJUSTMENT FACTORS

Name	Min.	Mean	Max.	Std. dev.	Distribution
Data communications	2	3.76	5	0.75	(Normal)
Distributed data processing	3	4.06	5	0.66	(Normal)
Performance	4	4.71	5	0.47	(Normal)
Heavily used configuration	2	2.65	4	0.70	(Normal)
Transaction rate	4	4.24	5	0.44	(Normal)
Online data entry	4	4.29	5	0.47	(Normal)
End user efficiency	4	4.24	5	0.44	(Normal)
Online update	3	3.35	5	0.70	(Normal)
Complex processing	4	4.35	5	0.49	(Normal)
Reusability	2	2.47	4	0.72	(Normal)
Installation ease	2	2.18	3	0.39	(Normal)
Operational ease	3	3.47	5	0.72	(Normal)
Multiple sites	3	3.29	5	0.59	(Normal)
Facilitate change	1	1.47	3	0.72	(Normal)

TABLE IV
RESULT OF FINAL, TRADITIONAL, AND STOCHASTIC FP

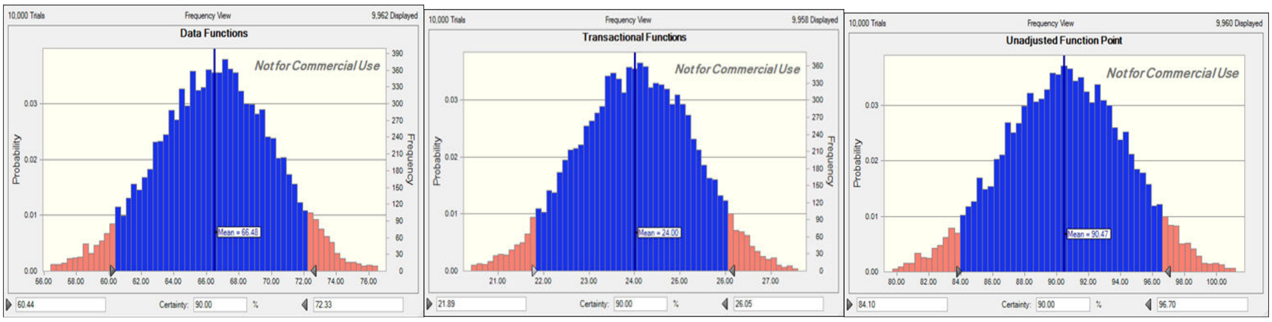
Component	Final	Traditional	Stochastic
Vehicle	141.95	157.14	134.14
Engine Torque	106.25	113.20	104.66
Air	96.18	112.14	84.14
Fuel	137.12	132.11	135.15
Ignition	129.82	134.14	109.15
Exhaust	109.28	124.21	102.14
Monitoring	52.76	65.47	54.87
Security	62.18	71.14	64.14

TABLE V
PAIRED T-TEST OF FINAL, TRADITIONAL, AND STOCHASTIC FP

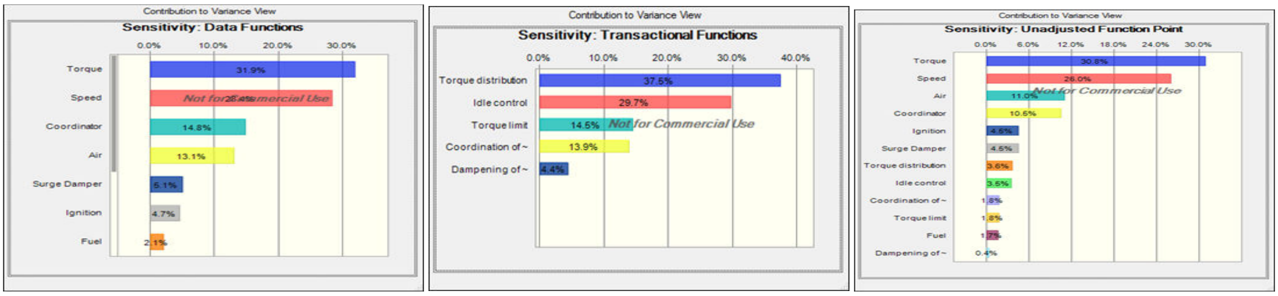
Division	Final	Traditional	Final	Stochastic
Mean	104.44	113.69	104.44	98.55
Std. dev.	33.10	31.38	33.10	29.44
Sample	8	8	8	8
t-value	3.366		-2.150	
Sig. Pro.	.008		.069	

2) LOC Correlation Comparison Verification

In terms of data collected in Table VI, we evaluated them indirectly by comparing the correlation between LOC and function point measured by traditional method and the correlation between LOC and function point measured by the way we suggested.

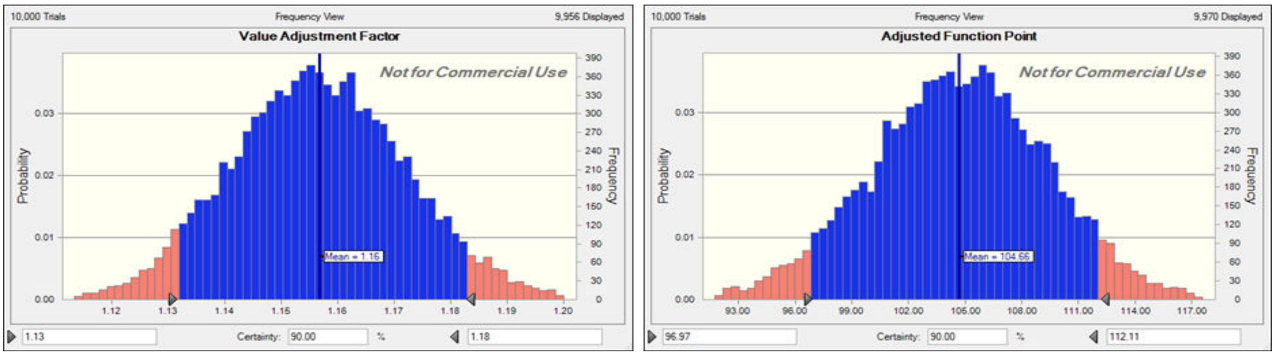


(a) Simulation Result of Data Function (b) Simulation Result of Transaction Function (c) Simulation Result of Unadjusted Function Point

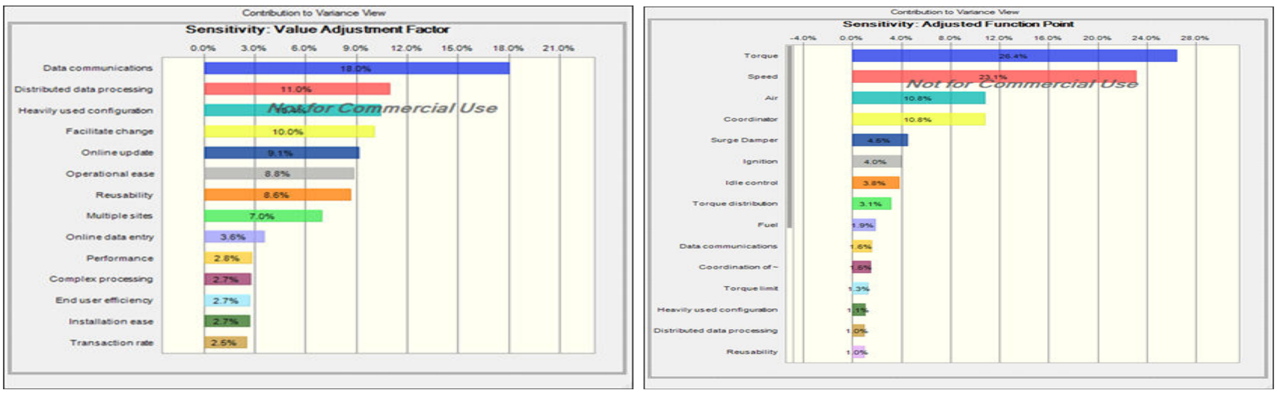


(d) Sensitive Analysis Result of Data Function(e) Sensitive Analysis Result of Transaction Function(f) Sensitivity Analysis Result of Unadjusted FP

Fig. 6 Monte Carlo Simulation Result of Data, Transaction Function and Unadjusted Function Point



(a) Simulation Result of Value Adjustment Factors (b) Simulation Result of Adjusted Function Point



(c) Sensitive Analysis Result of Value Adjustment Factors (d) Sensitive Analysis Result of Adjusted Function Point

Fig. 7 Monte Carlo Simulation Result of Value Adjustment Factors and Adjusted Function Point

The correlation coefficient between LOC and function point measured by traditional method is .737 and the correlation coefficient between LOC and function point measured by the way using stochastic distributions is .787 in the significance level .05 as described in Table VII. It is confirmed that both correlation coefficient are high, but correlation coefficient between LOC and function point measured by the way using stochastic distribution is slightly higher. However, LOC Correlation comparison is unreasonable to generalize as the data used only in the present study.

TABLE VI
LOC AND RESULT OF TRADITIONAL AND STOCHASTIC FP

Component	Final	Traditional	Stochastic
Vehicle	12,846	157.14	134.14
Engine Torque	15,284	113.20	104.66
Air	13,548	112.14	84.14
Fuel	18,246	132.11	135.15
Ignition	17,168	134.14	109.15
Exhaust	10,846	124.21	102.14
Monitoring	7,641	65.47	54.87
Security	6,521	71.14	64.14

TABLE VII
CORRELATION COEFFICIENT (R^2) BETWEEN LOC AND TRADITIONAL AND STOCHASTIC FP

Division	Traditional	Stochastic	Sig. Level
LOC	.737*	.787*	*p < .05

V. CONCLUSION AND FUTURE WORKS

In this work, we proposed a process for estimating function point using stochastic distribution. We applied our approach for an automotive E/E software development and demonstrated comparing our method with original process through statistical analysis. Effects of the study are as follows:

First, project managers can perform correct decisions based on the stochastic distribution in consideration of the variation in the process. It is difficult to understand numbers of variations that may occur at the beginning of the project lifecycle. Therefore, it is possible to increase the accuracy of the estimation at the beginning of the project, by defining a stochastic distribution for the process variation factors.

Second, project managers may be able to prepare the various alternatives in advance through simulation. The result of function point affects the effort and cost. If function point is high or low as a result of the simulation, it can be used as a base material for proper effort and cost adjustments. In addition, through sensitivity analysis project manager would identify significant factors to the measurement result and they are used as an object of the priority control in implementation and verification.

Third, project managers can reduce effort in estimating projects size and cost. In organizations that develop domain-specific similar systems such as automotive E/E control areas it can improve efficiency in measurement by using stochastic distribution derived from past projects.

In order to increase the completeness of the results of this study, it is necessary to expand and validate projects in various domains such as body, chassis, and multimedia. In addition, the result of function point is meaningful when used as monitoring indicators during project period and when project managers estimate effort and cost. This will extend the study of quantitative project management system based on the stochastic distribution associated with the EVM (Earned Value Management) based on function point.

REFERENCES

[1] U. S. Rao, K. Srikanth and P. Chinmay, "Stochastic Optimization Modeling and Quantitative Project Management", *IEEE Software*, 2008.
[2] M. K. Khedr, "Project risk management using Monte Carlo simulation," *AACE International Transactions*, 2006.
[3] B. Ergashev, "Estimating the lognormal-gamma model of operational risk using the Markov chain Monte Carlo method," Available at SSRN1316428, 2009.
[4] J. Dasgupta, G. Sahoo and R. P. Mohanty, "Monte carlo simulation based estimations: Case from a global outsourcing company," in *Technology Management Conference (ITMC), IEEE International, IEEE*, 2011.
[5] CMMI Product Team, "CMMI for Development Version 1.3," *Software Engineering Institute of Carnegie Mellon*, 2010.
[6] C. Y. Kim and H. S. Han, "Applying Monte Carlo Simulation for Software Project Risk Management Method," MA. *dissertation, University of SangMyung*, Seoul, Korea, 2011.
[7] P. Musilek, W. Pedrycz, N. Sun and G. Succi, "On the sensitivity of the COCOMO II Software Cost Estimation model," in *Proceedings of the eighth symposium on Software Metrics, IEEE Computer Society*, 2002.
[8] S. G. Park and J. Y. Park, "A Study for Software Sizing Method," *The Korea Computer Industry Education Society*, vol. 2, pp 471-480, 2004.
[9] J. S. Srivastava and G. Singh, "Optimized GSCs in Function Point Analysis - A Modified Approach," *International Journal of Research and Reviews in Applied Sciences*, vol. 17, 2013.
[10] H. K. Raju, Y. T. Krishnegowda, "Software Sizing and Productivity with Function Points," *Lecture Notes on Software Engineering*, vol. 1, 2013.