University of Central Florida

STARS

Electronic Theses and Dissertations

2018

A Methodology on Weapon Combat Effectiveness Analytics using Big Data and Live, Virtual, or/and Constructive Simulations

Won Jung II
University of Central Florida

Part of the Industrial Engineering Commons
Find similar works at: https://stars.library.ucf.edu/etd
University of Central Florida Libraries http://library.ucf.edu

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Jung, Won II, "A Methodology on Weapon Combat Effectiveness Analytics using Big Data and Live, Virtual, or/and Constructive Simulations" (2018). *Electronic Theses and Dissertations*. 5889.
https://stars.library.ucf.edu/etd/5889

A METHODOLOGY ON WEAPON COMBAT EFFECTIVENESS ANALYTICS USING BIG DATA AND LIVE, VIRTUAL, OR/AND CONSTRUCTIVE SIMULATIONS

by

WON IL JUNG
B.E. Korea Military Academy, 2003
B.S. Korea National Institute for Lifelong Education, 2007
M.S. Korea National Defense University, 2011
M.S. University of Central Florida, 2017

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2018

Major Professor: Gene Lee

© 2018 Won Il Jung

ABSTRACT

The Weapon Combat Effectiveness (WCE) analytics is very expensive, time-consuming, and dangerous in the real world because we have to create data from the real operations with a lot of people and weapons in the actual environment. The Modeling and Simulation (M&S) of many techniques is used for overcoming these limitations. Although the era of big data has emerged and achieved a great deal of success in a variety of fields, most WCE research using the Defense Modeling and Simulation (DM&S) techniques were studied without the help of big data technologies and techniques. The existing research has not considered various factors affecting WCE. This is because current research has been restricted by only using constructive simulation, a single weapon system, and limited scenarios. Therefore, the WCE analytics using existing methodologies have also incorporated the same limitations, and therefore, cannot help but get biased results.

To solve the above problem, this dissertation is to initially review and compose the basic knowledge for the new WCE analytics methodology using big data and DM&S to further serve as the stepping-stone of the future research for the interested researchers. Also, this dissertation presents the new methodology on WCE analytics using big data generated by Live, Virtual, or/and Constructive (LVC) simulations. This methodology can increase the fidelity of WCE analytics results by considering various factors. It can give opportunities for application of weapon acquisition, operations analytics and plan, and objective level development on each training factor for the weapon operators according to the selection of Measures of Effectiveness (MOEs) and Measures of Performance (MOPs), or impact factors, based on the analytics goal.

This work is dedicated to my wife

Without her support, this work would have not been possible.

ACKNOWLEDGMENTS

I would like to express my profound gratitude to many people who have guided and supported this research work and my life.

I would particularly like to thank my supervisors, Dr. Gene H. Lee and Dr. Luis C. Rabelo for the good ideas and inspiration from the conception to completion of this dissertation. Appreciation is also extended to my dissertation committee members, Dr. Ahmad Elshennawy and Dr. Ali Ahmad for their valuable time and contributions.

I also thank Dr. William J. Thompson, Madelda Thompson, Mingyu Cha, and Larry Lowe who have helped me with this research in many ways. I'd like to express my gratitude to the LAB members, Kyungeun Lee, Donghee Baek, Jaeho Kim, and Dr. Mario Marin.

My deepest gratitude goes to my father, Pilwha Jung, and my mother, Jaeeun Kim, who have always proactively supported me through all areas in life. Also, I would like to express my gratitude to my father-in-law, Pilgyo Jung, and my mother-in-law, Sunye Lee.

I also extend my thanks to my adorable 10-year-old son, Hayun, and 7-year-old daughter, Yunseo, who have been so very patient and willing to help me in any way. Most importantly, I greatly thank my wife, Euichun, for loving and supporting me during this process. She has never faltered and been strong during this time.

First and foremost, I thank my almighty God for His grace and unconditional love and for all He does for me.

LVC Interoperability Laboratory, Orlando, FL., USA

March 6th, 2018

Won Il Jung

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

6DoF	Six Degrees of Freedom
AAsim	Army Aviation simulation
ACMI	Air Combat Maneuvering Instrumentation
ADC	Availability, Dependability, and Capability
AFSIM	Advanced Framework for Simulation, Integration and Modeling
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
AIC	Akaike information criterion
ALSP	Aggregate Level Simulation Protocol
AMG	Architecture Management Group
AMSAA	Army Materiel Systems Analysis Activity
AMT	Architecture Management Team
ANOVA	ANalysis Of VAriance
ANP	Analytical Net Process
API	Application Programmers Interface
ASW	Anti-Submarine Warfare
BDP	Big Data Preparation
BODP	Biaxial Optical Detection Platform
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance

CALL	Center for Army Lessons Learned
CATT	Combined Arms Tactical Training System
CE	Combat Effectiveness
CERs	Cost Estimation Relationships
CGF	Computer Generated Force
COTS	Commercial Off-The-Shelf
CPU	Central Processing Unit
CTIA	Common Training Instrumentation Architecture
CTIS	Combat Training Instrumentation Systems
DARPA	Defense Advanced Research Projects Agency
DBD	Defense Big Data
DES	Discrete Event System
DEVS	Discrete Event Systems Specification
DFS	Distributed File System
DIS	Distributed Interactive Simulation
DM&S	Defense Modeling and Simulation
DMSO	Defense Modeling and Simulation Office
DoD	Department of Defense
GOTS	GOvernment off The Shelf
GPS	Global Positioning System
GUI	Graphical User Interface
HDD	Hard Disk Drive

HDFS	Hadoop Distributed File Systems
HLA	High Level Architecture
IEEE	Institute of Electrical and Electronics Engineers
IDC	International Data Corporation
IWG	Interface Working Group
JHMCS	Joint Helmet Mounted Cueing System
JMEMs	Joint Munitions Effectiveness Manuals
JSAF	Joint Semi-Automated Forces
JTCG/ME	Joint Technical Coordinating Group for Munitions Effectiveness
JVM	Java Virtual Machine
KCTC	Korea Combat Training Center
LC	Live and Constructive
LROM	Logical Range Object Model
LT2	Live Training Transformation
LV	Live and Virtual
LVC	Live, Virtual, or/and Constructive
LVCAR	Live Virtual Constructive Architecture Roadmap
M&S	Modeling and Simulation
MILES	Multiple Integrated Laser Engagement System
MIP	Mixed Integer Programming
MMRE	Mean Magnitude of Relative Error
MOEs	Measures of Effectiveness

MOPs	Measures of Performance
MW	Middleware
OM	Object Model
OMT	Object Model Template
OneSAF	One Semi-Automated Forces
OS	Operating System
PDU	Protocol Data Unit
Peo-Stri	Program Executive Office for Simulation, Training, and Instrumentation
PLAF	Product Line Architecture Framework
PRED	Prediction
QFD	Quality Function Deployment
RAM	Random-Access Memory
RCS	Radar Cross Section
ROC	Required Operational Capability
RTI	Run-time Infrastructure
SACM	Small Advanced Capability Missile
SAF	Semi-Automated Forces
SDA	Software Development Activity
SGI	Silicon Graphics Inc.
SIMNET	Simulation Networking
SISO	Simulation Interoperability Standards
SMEs	Subject-Matter Experts

SSAs	Standard Simulation Architectures
STORM	Synthetic Theater Operations Research Model
TENA	Training Enabling Architecture
UAPs	Unified Action Partners
UAV	Unmanned Aerial Vehicles
UGV	Unmanned Ground Vehicle
V&V	Validation &Verification
VC	Virtual and Constructive
VGA	Video Graphic Array
VIF	Variance Inflation Factor
WCE	Weapon Combat Effectiveness
WCEEs	Weapon Combat Effectiveness Equations
WSoS	Weapon System-of-System
WTA	Weapon Target Allocation

CHAPTER ONE: INTRODUCTION

This chapter explains the background related to WCE analytics. Problems, which lead to treacherous results on the existing WCE analytics, are identified and stated. The approach to solve the problems is suggested. The purpose, goal, objectives are stated to guide this research range and direction and achieve the approach. The potential contributions are shown on the new WCE analytics methodology using big data and simulations.

1.1 Background

The victory or defeat in war is a matter of life and death for the nation. Therefore, the nation must prepare and prevent war against realistic and potential enemies by keeping and improving the national power continuously for insurance. The evaluation on the national power is an important process to check the war preparation status and improve it effectively and efficiently. At this moment, since the effectiveness analytics of the military force is directly related to war, it is the critical factor of the evaluation on the national power. Considering Combat Effectiveness (CE) is the measurement of the ability of a military force to achieve the military objective in a combat operation using limited resources (Hodgson, 1957; Zhao, Yin, & Song, 2016), CE is an essential criterion to evaluate the military force. Particularly, there is no doubt that Weapon Combat Effectiveness (WCE) of CE is the key part in the modern and future wars. This is because the effectiveness based weapon systems are more important in the war as the paradigm is transformed from the linear to the non-linear war, which is mainly executed by the guerilla strategy with the special forces (Hammes, 2004), such as warfare in Ukraine, Afghanistan, and Iraq. At the moment, WCE can be defined as the performance measurement of the weapon system needed for a military force to affectively complete the given mission in a combat operation. Figure 1 shows the necessity of WCE analytics.

Nowadays, in order to protect their own citizens and territory, most nations are devoting portions of their budgets to developing and purchasing the newest weapon systems based on the WCE (Perlo-Freeman, Fleurant, Wezeman, & Wezeman, 2016). Nations are attempting to improve methodologies on the operations plan and analytics to enable leaders to make tactical or strategic decisions, such as the weapon allocation and the weapon system's exploitation, by using WCE (W. Jung, Shin, Mohamed, Rabelo, & Lee, 2016).

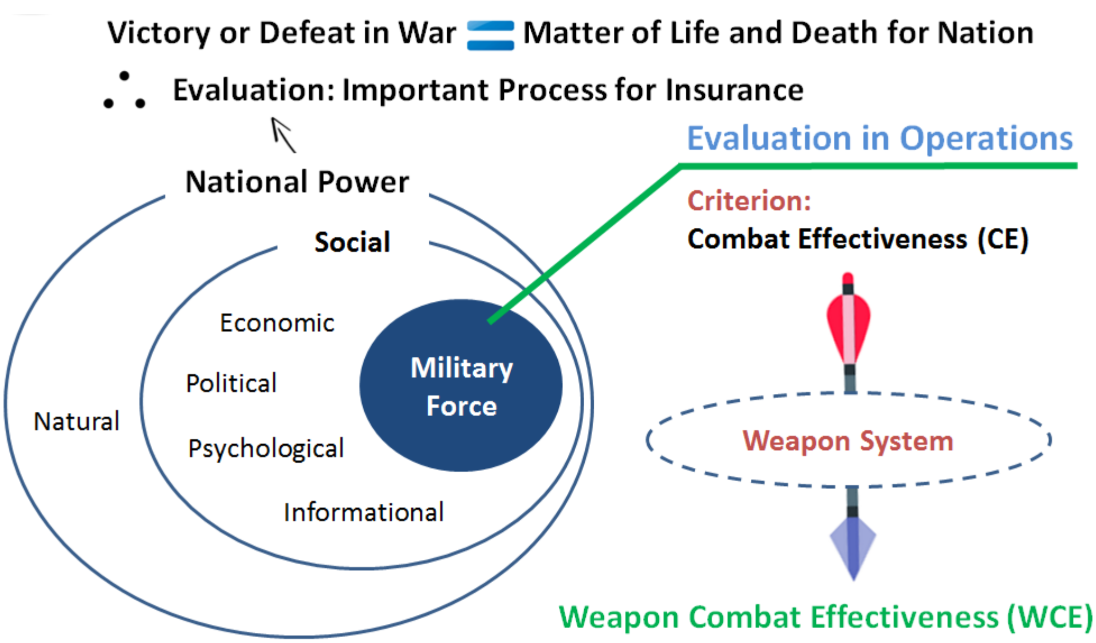

Figure 1. Necessity of WCE Analytics

Even though WCE analytics is such an important process, the Department of Defense (DoD) must operate under the assumption of budget restrictions because methods for savings in the defense budget must be found continuously. Furthermore, WCE analytics is demanding work to concisely measure the WCE due to the complexity of the combat environment. As a matter of fact, we must analyze WCE under diverse situations in order to reach the best conclusion. However, if the WCE analytics with a variety of scenarios is executed in the real world, it is not only a very costly, time-consuming task, but also a dangerous one. Deployment and operations of all weapon systems for WCE analytics is limited in real environment. Weapon systems are becoming more complex and diverse so that the WCE analytics can be difficult to use in real environments, where time and space are finite. Moreover, the experimental testing in real environments could cause diplomatic challenges from neighboring countries. Therefore, many nations consider a variety of methods to solve the above problems as economically and precisely as possible to analyze the WCE.

Approximately fifty percent of the papers related to WCE based on the Compendex Database applied the most popular method, Defense Modeling and Simulation (DM&S), to analyze WCE. This is because Modeling and Simulation (M&S) enables researchers to reduce cost, time, and risk of experiment in the real world (Schmidt, 1978). Limitations on weapon deployment and operations in real environment can be solved using M&S. Complex and various weapon systems can use M&S techniques in virtual environment. Although researchers utilize DM&S to effectively analyze the WCE, they cannot identify important factors and the possibility of overlooking significant factors that really affect WCE. Therefore, the results could be biased under the above problems (W. Jung et al., 2016). Figure 2 shows problems of WCE analytics in real world and the most popular method for overcoming the problems.

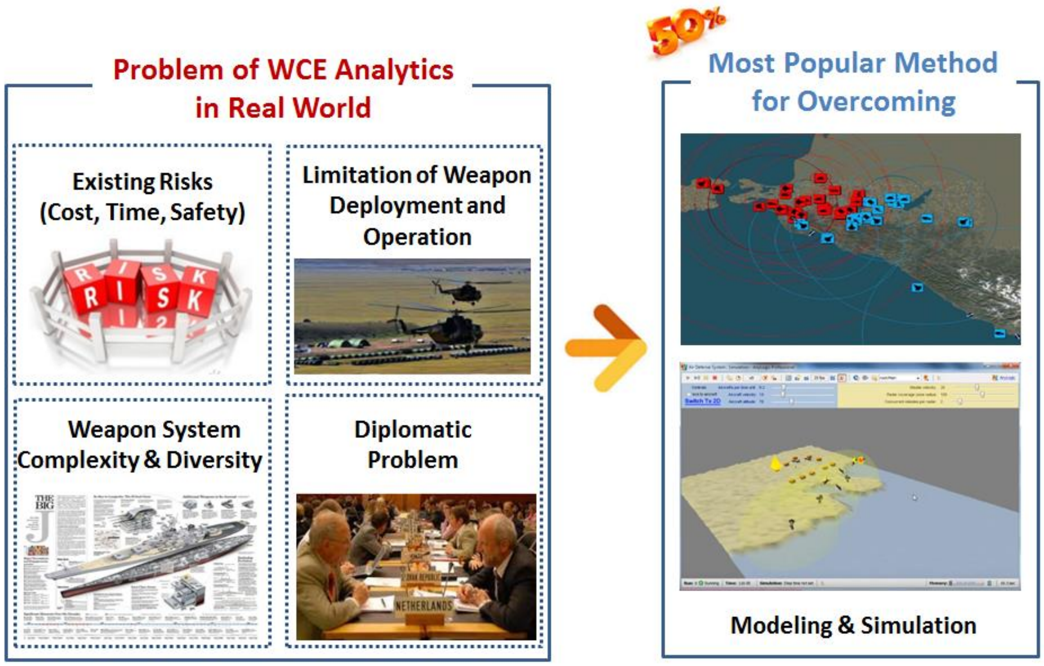

Figure 2. Problems of WCE Analytics in Real World and Most Popular Solution Tool

Big data has transformed the traditional point of view in science and engineering (Shi & Abdel-Aty, 2015). Before the era of big data, huge amounts of data could not be stored and complete results not produced within a proper time frame (Khan et al., 2014). However, big data advanced techniques have contributed to enabling us to store and process huge amounts of data as well as a variety of data within an expected time frame. Currently, the USA DoD invests 250 million dollars every year for big data analytics as well as scientific research that support the defense strategy using these techniques (Moorthy et al., 2015). This means that big data has proven itself in the area of defense.

1.2 Problem Statement

Until recently WCE analytics research did not satisfy or try to satisfy two main principles of sampling, which are the statistical regularity and inertia of large numbers. Therefore, the WCE analytics results have not had a valid statistical inference regarding a sampling population without satisfying the above principles of the statistical and inertia of large numbers (Sharma, 2012). This proves that the analytics results are insufficient to explain statistically, and thus, cannot be trusted entirely. This problem comes from not considering a variety of factors influencing the WCE because the previous studies were limited to very basic constructive simulations and a single weapon system with their own created scenarios. This means that the WCE results were analyzed based on small and biased samples, which ignore the statistical regularity. Also, although advanced techniques and technologies, which are related to big data and integration of distributed systems, are being developed, the existing work has not accepted the challenge of including the principle of inertia of large numbers in the research. It is to miss the opportunity to improve the fidelity of the WCE analytics. Figure 3 summarizes the identified problem.

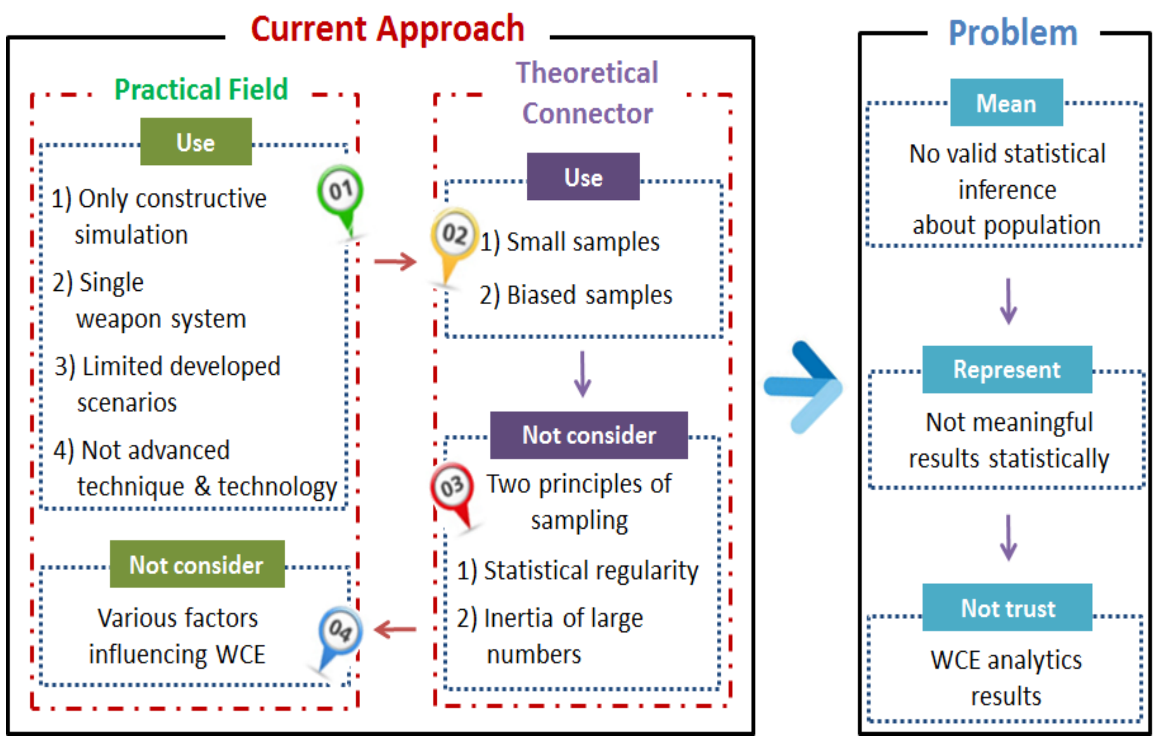

Figure 3. Summary of Problem Statement

1.3 Proposed Approach

Big data can be generated through the LVC simulation model federated by Standard Simulation Architectures (SSAs). This is because the simulation model can more represent real-world circumstances rather than the use of a single constructive simulation or connected constructive simulations utilizing advantages of heterogeneous distributed models. Virtual simulation can overcome limitations of constructive simulations by considering human factors which exist between the human and the weapon system. If a live simulation is applied to big data generation to analyze WCE, the generated data can reflect the real environment. The advanced computing techniques and technologies can support the ability to consider and analyze various factors based on the big data generated by LVC simulations under various scenarios. Also, the data accumulated and analyzed by existing models can be used to analyze WCE by integrating with big data generated by LVC simulations. This means that the WCE can be analyzed based on the samples similar to a population rather than the existing methods when using the suggested new approach. This approach enables the results of the WCE analytics to have high fidelity by satisfying two main principles of sampling. The new methodology on the WCE analytics using big data generated by LVC simulations is needed for solving the identified problems. Figure 4 shows the new approach for solving the problems. Although the amount of research related to WEC, DM&S, and big data generated to date is large, there is not any research related to the intersection of WCE, DM&S, and big data as shown in Figure 5. Therefore, this approach is meaningful as the first WCE analytics research using DM&S and big data.

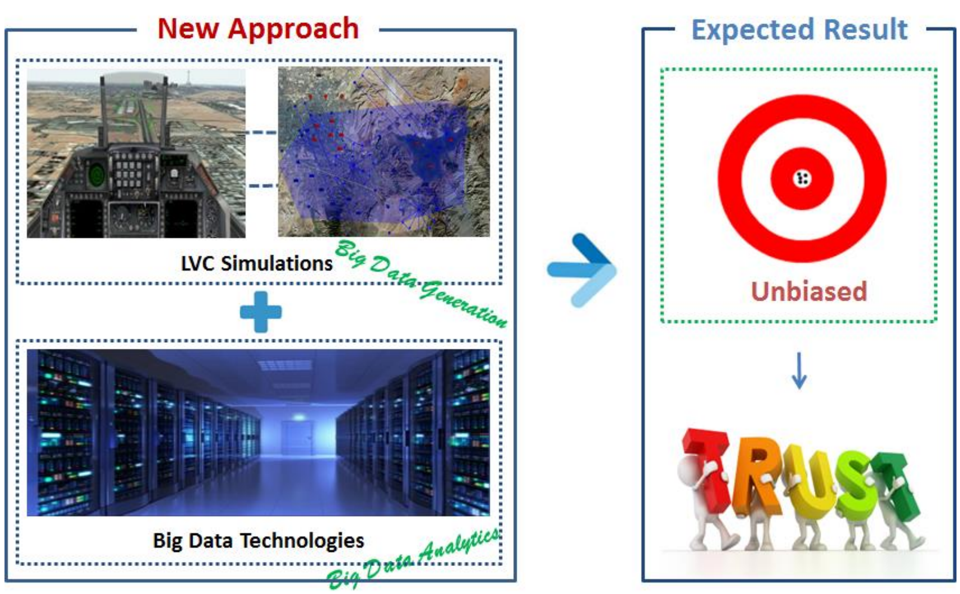

Figure 4. Proposed Approach for Solving Identified Problems from Existing Research

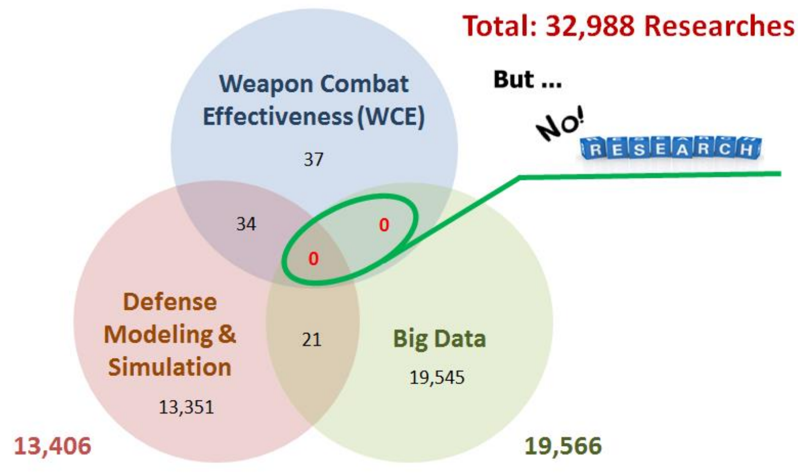

Figure 5. The Number of Research related to WEC, DM&S, or Big Data based on the Compendex Database (Sep. 21, 2016)

1.4 Purpose, Goal, and Objectives of the Research

This part identifies the purpose, goal, and objectives of this research. These enable us to limit the research range, guide the research direction, and achieve the above proposed approach. This research's purpose is to provide higher accuracy of WCE analytics which could lead to more effective and precise decision making in the defense field. For achieving the purpose, this research's goal is to develop a new methodology on WCE analytics using big data generated by LVC simulations. There are nine objectives to complete the goal. The first objective is to classify existing research related to WCE analytics which will enable us to structurally understand existing research and provide the information for developing a new methodology. The second one is to review DM&S and the big data research. This objective helps us to understand the latest techniques and technologies of DM&S and big data as well as to identify applicable techniques and technologies for a new methodology. The third is to evaluate existing research on WCE analytics using M&S. The direction to develop the new methodology is guided by the third objective where research gaps can be identified. The fourth objective is to construct the system for VC simulations that must be validated and verified. This objective is to show how to build an environment in which a new methodology is applied for generating big data on WCE. A live simulation is conceptually connected to VC simulations but physically disconnected in this research. The fifth one represents a way to build the system for analytics of big data generated by VC simulations. This is necessary to apply a new methodology as a part of big data analytics. The sixth objective is to develop an algorithm used in the new methodology. The algorithm's job is to estimate the relationship between the WCE variable and independent variables which influence WCE. The seventh objective is to create visual formats for representing the WCE analytics results. This enables decision makers to effectively and efficiently identify and decide the best alternative. In the eighth, a case study is conducted in support of showing how to solve a problem using the new methodology and its practical benefits. The ninth objective is to suggest areas of extension where the new methodology can be beneficial. This part shows the future utilitarian value of the new methodology, which has the potential to improve the accuracy of analytics in difficult environments. These environments make it hard to acquire a variety of real data. Table 1 summarizes the purpose, goal, and objectives of this research.

Table 1. Summary of Purpose, Goal, and Objectives of this Research

Purpose	Improve the accuracy of WCE analytics
Goal	Suggest the new methodology on WCE analytics using LVC simulations and big data
Objectives (9)	1) Classification of existing research on WCE analytics
	2) Review of research related to DM&S and big data
	3) Assessment of existing research related to WCE analytics using M&S
	4) Build an environment for VC simulations to generate big data on WCE (including a limited live simulation)
	5) Develop an environment for big data analytics on WCE
	6) Build an algorithm for Weapon Combat Effectiveness Equation (WCEE) development
	7) Develop formats for visualization of WCE analytics results
	8) Case study using the new methodology on WCE analytics
	9) Suggest extendable areas to use the new methodology on WCE analytics

1.5 Potential Contribution

The contribution of this paper is to initially collect the basic knowledge for the new WCE analytics methodology using big data and DM&S to further serve as the stepping-stone of the future research. The achievements for this are shown in the following: (a) general overview of WCE, DM&S, and Defense Big Data (DBD), (b) establishment of WCE classification, (c) identification of the research challenges, and (d) investigation and suggestion of solutions for overcoming research challenges.

Furthermore, a new methodology for the WCE analytics is proposed. The method for WCE analytics is a new concept that enables us to have results of higher fidelity by considering and analyzing a variety of variables required to evaluate weapon effectiveness under various scenarios. This can be achieved because, from utilizing big data techniques and technologies as well as LVC simulations, the new method includes the following benefits: (a) utility of abundant data, a-1) LVC simulations benefits utility, a-2) big data technique benefits utility, and a-3) external source utility, (b) modeling reality; b-1) assumption degree minimization, b-2) various factors application, and b-3) various scenario application, (c) generalization; c-1) application flexibility, and c-2) comprehensive analysis, (d) analytics result usability; d-1) WCE equation estimation with various factors, and d-2) optimal values recommendation based on the constraints.

Lastly, this paper shows that the suggested methodology has the possibility of expansion into various fields, such as weapon acquisition, and operations analytics and plan. Also, another expansion area is the objective level development on each training factor for the weapon operators according to the selection of Measures of Effectiveness (MOEs) and Measures of Performance (MOPs) based on the analytics goal. Figure 6 summarizes the potential contribution of this paper.

Figure 6. Potential Contribution of New Methodology

1.6 Synopsis

This remaining section of the dissertation is organized as follows. Chapter 2 represents a general review of research related to WCE. This chapter includes WCE classification and discussion of existing papers of WCE based on the classification. Also, reviewed is DM&S related to big data generation for WCE analytics. The critical issues are reviewed for general big data followed by DBD. Lastly, the research gaps are analyzed based on existing research on WCE analytics using M&S. Chapter 3 introduces a new methodology on WCE analytics using big data generated by LVC simulations. This chapter explains the new methodology divided into five steps, which are generating big data, followed by collecting additional data related to WCE, processing big data for WCE analytics, estimating WCE equation and optimal values, and reporting results. Chapter 4 represents the case study using the new methodology on WCE analytics. This case study is done according to the five steps of the new methodology. Chapter 5 reveals three extendable areas which are the weapon acquisition, the operations plan and analytics, and the objective level development for each operator's training factor using the new methodology within the defense field. Chapter 6 concludes this dissertation. The chapter summarizes this research, underlines its contributions, and suggests guidelines for future research by identifying its limitations.

CHAPTER TWO: LITERATURE REVIEW

This chapter reviews literature related to the WCE analytics. The literature studied on the WCE analytics in this dissertation is classified into three viewpoints which are reliability, efficiency, and economics. Also, the present status of DM&S and big data is reviewed to obtain background knowledge for the development of the new WCE analytics methodology. Finally, research gaps are analyzed based on the existing research.

2.1 Weapon Combat Effectiveness Classification

There is much research related to WCE for effectively strengthening the power of defense under a limited budget. Particularly, researchers have mainly focused on studying WCE to improve reliability, efficiency, and economics. They have confronted and solved real world problems by applying the results of current research. In this section, WCE research is classified as shown in Figure 7 and Table 2 and then reviewed according to the classification.

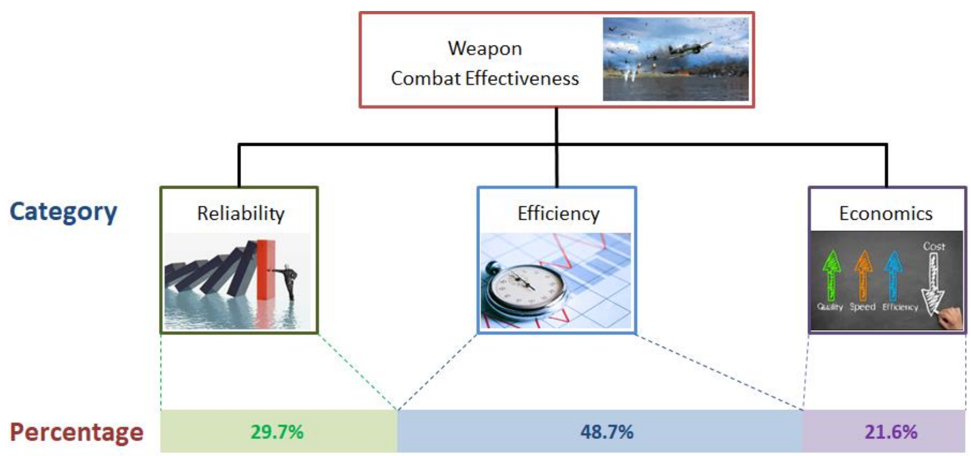

Figure 7. Research Percentage per a Type of WCE Classification

Table 2. Classification of WCE Studied according to Analytics Purpose

Classification		Authors (Year)	Weapon (System)
Reliability	Improving fidelity	Peng et al. (2015)	Submarine-to-air missile
	Mitigating uncertainty	Li et al. (2007) Zheng (2011)	Weapon System-of-System Radar electronic warfare
	Analyzing from more based part	Qi et al. (2011) Pan et al. (2006) Seo et al. (2011) Jin et al. (2014) Skow (1992) Liu et al. (2013)	Missile defense system Air-defense weapons Anti-torpedo defense system General system Aircraft Air-to-ground weapon
	Guaranteeing data accuracy	Zhou et al. (2015) Song (2015)	Airborne laser weapon Anti-ship missile
Efficiency	Mitigating complexity	Liu et al. (2011)	Anti-radiation weapon
	Suggesting comprehensive method	Le et al. (2013) Liou et al. (2014) Liou et al. (2008) Tong et al. (2007) Chusilp et al. (2014) Wu et al. (2014)	Weapon System-of-System Weapon System-of-System Weapon System-of-System General system Missile system General system
	Optimizing weapon allocation	Lee and Lo (1994) Ruan et al. (2010)	Air defense Fleet antiaircraft
	Efficiently operating weapons	Zheng et al. (2009) Li et al. (2015) Chio et al. (2010)	Air combat Vehicle UAV
	Establishing relationship between weapon character and effectiveness	Boppe et al. (1994) Jette (2000) Yan et al. (2007)	Missile Missile Missile
	Standardizing the effectiveness	Ji et al. (2005) Wang et al. (2007)	Ground weapon Anti-ship combat
	Integrating systems	Osder (1991)	Attack helicopter
	Developing or improving weapon systems	Boppe et al. (1994) Jette (2000) Tuttle (2003) M. J. (2005)	Aircraft combat Soldier combat Aircraft Aircraft
Economics	-	Ludvik et al. (1996) Lim et al. (2009) Jung et al. (2012) Hopkins et al. (1966) Kim (2013) Yin et al. (2015) Liang et al. (2006) Ru et al. (2012)	Rocket weapon Rocket weapon Artillery weapon Surveillance device Helicopter General system Torpedo weapon General

2.1.1 Reliability

The reliability of the weapon system and its evaluation methods are very important factors to help maximize the WCE. The efforts to improve the reliability have been shown through much research. Specifically, researchers have focused on a) improving fidelity, b) mitigating uncertainty, c) analyzing from more based parts, and d) guaranteeing data accuracy. Researching to achieve the reliability of WCE through the fidelity enhancement, Peng, Wang, and Zou (2015) built a new model developed by combining a grey cloud model and Analytic Hierarchy Process (AHP). The developed model improved the fidelity and obtained the reliability of combat effectiveness evaluation for submarine-to-air missile. Some research paid attention to the achievement of the reliability of WCE by mitigating the uncertainty. X. Li, Tan, and Yang (2007) developed the effectiveness evaluation method based on exploratory analytics, which handles the uncertainty of the battle environment, for Weapon System-of-System (WSoS) using the multi-resolution model by combining simulation and analytical modeling. Y. Zheng (2011) applied the Bayesian Network to the combat effectiveness evaluation of the radar electronic warfare. By applying the Bayesian Network concept, complex relationship and uncertainty of probability could be expressed in the analysis model. Some researchers considered more detailed analytics to obtain the reliability of WCE. Qi, Liu, Liang, and Niu (2011) suggested the effective fuzzy evaluation model on missile defense system effectiveness. The weight coefficients of the main factors affecting the defense system were given using the fuzzy optimization algorithm. Pan, Ma, and Li (2006) analyzed a variety of air-defense weapons combat effectiveness using the queueing network method. This research contributed to the ease in the analysis of the combat effectiveness for each component. Seo, Song, Kwon, and Kim (2011) analyzed the effectiveness of a battle ship's anti-torpedo defense system by varying tactics and weapon performance using the Discrete Event Systems Specification (DEVS) formalism. Jin, Pang, Li, Yuan, and Wu (2014) analyzed the relationship between human factors and ability of performance, which is an important factor for increasing the combat effectiveness in regard to the human-system combination, and studied the improvement of job performance. Skow (1992) redefined the agility, which can be a significant factor to achieve the aircraft combat effectiveness, and suggested an analysis method for measuring all agility factors of the weapon system. C. Liu, Fan, Bao, and Wang (2013) designed the air-to-ground weapon launcher detector of a specific aircraft. The detector system improved the aircraft combat effectiveness by helping to more accurately detect missile launchers within a short time. There is research with efforts to enhance reliability of WCE by attaining basic accurate data. Zhou, Zhang, Zeng, and Zhang (2015) established the kill probability of the airborne laser weapon to an anti-missile defense system. The result was utilized for the improvement of the airborne laser WCE and detection of enemy airborne laser weapons. Song (2015) analyzed the penetration and damage capacity of an anti-ship missile with the resulting data having the capability of being used for tactical missile effectiveness estimation.

2.1.2 Efficiency

Researchers have focused on enhancing the efficiency of the WCE and its evaluation methods under its limited abilities. Much research has been done to achieve the efficiency in the following ways: a) mitigating complexity, b) suggesting comprehensive method, c) optimizing weapon allocation, d) operating weapons efficiently, e) identifying the relationship between the weapon characters and effectiveness f) standardizing the effectiveness, g) integrating systems, and h) developing or improving weapon systems.

In order to improve the efficiency of WCE with mitigating complexity, Y. Liu, Zhao, Wang, Wang, and Feng (2011) proposed a new method of the combat effectiveness evaluation using non-linear index aggregation for the anti-radiation weapon under complex environments.

Also, Liou, Huang, and Yang (2008), Liou, Cheng, Liou, and Liou (2014), and Le, Zhong, Jianqiang, and Xiongwei (2013) suggested a method to reduce the complexity of effectiveness evaluation indexes for WSoS using the autoencoder, which is an artificial neural network applied to obtain the unsupervised learning of proficient coding. Tong, Ye, Lu, and Hui (2007) proposed a model to improve problems of existing models, which are complex and theoretical to evaluate combat outcomes. Chusilp, Charubhun, and Koanantachai (2014) compared WCE by using a more complex P_k matrix and a simpler Carleton damage function. The Carleton damage function was recommended because there is a minor difference between the two ways in terms of analyzing WCE. As a way to improve efficiency of the WCE evaluation, Wu, Zhang, Zhang, and He (2014) suggested the fuzzy comprehensive evaluation method to adapt to the broad evaluation of the weapon equipment performance.

There is research to enhance WCE with the optimizing weapon assignment. Y. J. Lee and Lo (1994) suggested the air defense analytics method by applying inductive learning to simulation modeling. This method clearly discovered rules about the optimal weapon assignment. Ruan, Li, and Liu (2010) established a fleet antiaircraft Weapon Target Allocation

(WTA) model using the artificial immune algorithm. This study applied a chromosome coding of the special antibody data structure, which focused on the specified requirements for problem-solving and utilized the clone-immune operators for the effective allocation of weapons. Some parts of the research attempted to increase the efficiency of WCE by operating the weapons more efficiently. J. Zheng, Zhang, and Wang (2009) proposed a coordination method for selecting an optimal behavior in air combat using a multi-agent based model. The simulation results showed that the optimal behavior policy was the proposed validated model. K. Li, Wang, Lv, Gao, and Song (2015) proposed a new approach to find the perfect position and direction of a vehicle using Biaxial Optical Detection Platform (BODP). Chio, Weng, and Chang (2010) analyzed the Unmanned Aerial Vehicles (UAV) combat effectiveness by using a system dynamic model according to operational types.

Other research enhanced the efficiency of WCE by establishing the relationship between the weapon characters and effectiveness. Boppe et al. evaluated the missile combat effectiveness using the Availability, Dependability, and Capability (ADC) method. This paper showed that the Radar Cross Section (RCS) of the missiles has a positive relationship with detectable probability by the radar and the missile velocity has a negative relationship with the ability of the enemy's defense system (Boppe & Martorella, 1994; Jette, 2000; Yan, Gu, Guan, & Sun, 2007). In order to progress the efficiency of WCE evaluation, Ji, Liu, Wang, and Li (2005) built a method for suggesting the combat effectiveness standardization of ground weapons and calculated its values by using a case study. Also, Wang, Dong, Jiang, and Zhang (2007) analyzed the combination of parameters of the index system for anti-ship combat effectiveness of attack aircraft and established the sub-models for the combat effectiveness. As a way to achieve the efficiency of WCE, Osder (1991) improved an attack helicopter combat effectiveness by integrating fire and flight control systems which can enhance firing possibility, weapon accuracy, and survivability. Some research obtained the efficiency of WCE by developing or improving weapon systems. Boppe and Martorella (1994) improved aircraft combat effectiveness by overcoming the fighter maneuvering limitations through the controllability improvement with engine thrust-vectoring and thrust-reversing. Jette (2000) showed that infantry soldier combat effectiveness can be enhanced by integrating optics with displays and sensors in Land Warrior, which is the U.S. Army's premier program. Tuttle (2003) showed guided weapons using Global Positioning System (GPS) improved B-1B, B2, and B-52 combat effectiveness. M. J (2005) introduced the Joint Helmet Mounted Cueing System (JHMCS), which improved co-ordination and situation awareness, to guarantee combat effectiveness.

2.1.3 Economics

The acquisition cost of the weapon system has been increasing because weapon systems have become more precise and complex while using extremely advanced technology. Therefore, much research studied the cost based WCE to help decision makers select the optimal option. Ludvik and Konecny (1996) analyzed cost effectiveness on the rocket weapon system. Lim, Cho, and Park (2009) analyzed the combat effectiveness of the multiple launcher rocket based on the cost by using the AHP method. W. Jung, Lee, Kim, and Kang (2012) suggested the cost estimation methodology appropriate to the effectiveness of the artillery weapon system to be obtained in the future by using Cost Estimation Relationships (CERs) linear combinations.

Hopkins, Marr, Stachtchenko, and Baker (1966) analyzed the cost-effectiveness for the surveillance device, which supports short range surface-to-air missiles using the computer simulation. Kim (2013) analyzed the cost and effectiveness according to the survivability design level of the light armed helicopter to support the decision making using AHP, M&S, Quality Function Deployment (QFD), and parametric estimating method. The cost evaluation model for the weapon lifecycle can be established by evaluating the weapon effectiveness using Analytical Net Process (ANP) to obtain the optimal weapon system (Yin & Xie, 2015). Q. Liang, Song, and Pan (2006) studied the lifecycle cost and effectiveness of torpedo weapon systems using the fuzzy ideal point methodology. Ru and Gao (2012) suggested the criteria for WTA evaluation based on the cost and effectiveness of the weapons.

2.2 Defense Modeling and Simulation

This section explains the computer based DM&S related to big data generation for WCE analytics. The proper simulation models need to be selected in support of generating big data appropriate to WCE analytics. The selected simulation models can be models for live, virtual, and constructive simulations according to the analytics environment which includes the cost, the time, the technique level, and the required level. Also, the selected simulation models are required to be connected to other single simulation models with the interoperability protocol, which plays a role in the exchange of information without errors among the models. At this time, the Standard Simulation Architectures (SSAs) are generally developed and applied to achieve the interoperability among the selected simulation models which are standalone systems. Validation &Verification (V&V) processes are then needed to be executed to guarantee fidelity of the big data generated by the selected models.

2.2.1 Live, Virtual, and Constructive Simulations

This part explains the concepts of live, virtual, and constructive simulations and differentiates them based on the types of personnel, systems, and operation as shown in Figure 8. A live simulation is conducted by real people, real systems, and simulated operation. A virtual simulation is executed by real people, simulated systems, and simulated operation. A constructive simulation is done by simulated people, simulated systems, and simulated operation.

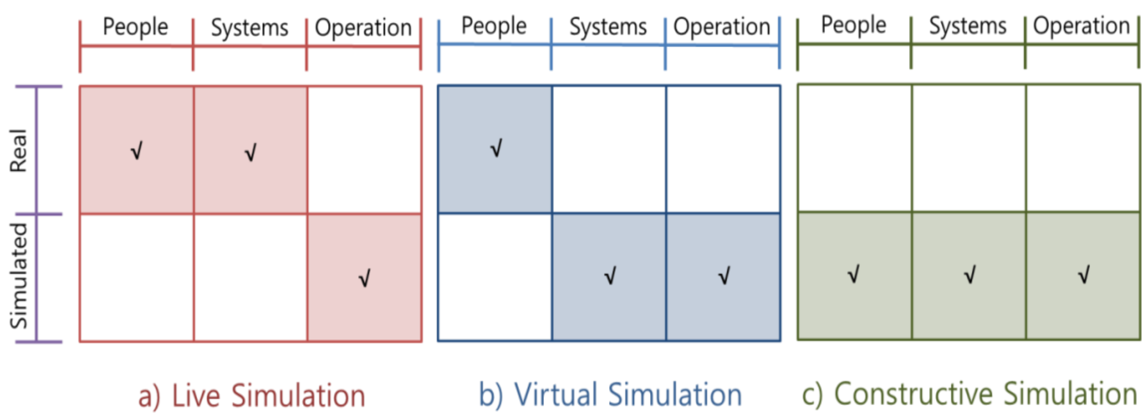

Figure 8. Classification of Live, Virtual, and Constructive Simulations

2.2.1.1 Live Simulation

A live simulation is when real people operate real systems to attend a simulated operation in a real environment (Tolk, 2012). Daly and Thorpe (2009) distinguished live simulation training from synthetic training, which is conducted with real people using real equipment and weapons in a virtual environment. The Air Combat Maneuvering Instrumentation (ACMI)

System, which is used by the air force for the air fighting training and analysis, is an example. The ACMI calculates and offers real aircraft information that includes time, speed, position, acceleration, orientation, and armed status. Also, the Korea Combat Training Center (KCTC), in which the engagement results between blue and red teams are computed by using computers, sensors, GPS, and a laser engagement system without exchanging real ammunition, is another example of a live simulation.

2.2.1.2 Virtual Simulation

When real people operate simulated systems to attend a simulated operation in a real environment, it is called a virtual simulation (Tolk, 2012). The virtual simulation reflects human cognitive and perceptual processes unlike a constructive simulation. It is a representative of a virtual simulation in which virtual operation war fighters use F-15 simulators that provides immediate feedback to the trainees. The fighters are immersed in a virtual environment, in which other weapon systems are displayed, such as aircrafts, tanks, and battle ships.

2.2.1.3 Constructive Simulation

It is a constructive simulation when simulated people operate simulated systems to attend a simulated operation (Tolk, 2012). Although real people compose scenarios for the simulations, they do not participate in determining the results. Also, the described range of the constructive simulation can be various according to the ability of the human operators who are limited to the number of simulated forces and the level of behaviors. The typical example is a war game, which focuses on military strategy, operations, or tactics.

2.2.2 Principal Levels of Military Models

Military models can be divided into four principal hierarchies, which consist of engineering, engagement, mission, and theatre, based on the force size, the objective, and resolution describing a simulation as shown in the Figure 9. Each hierarchy is explained below in more detail.

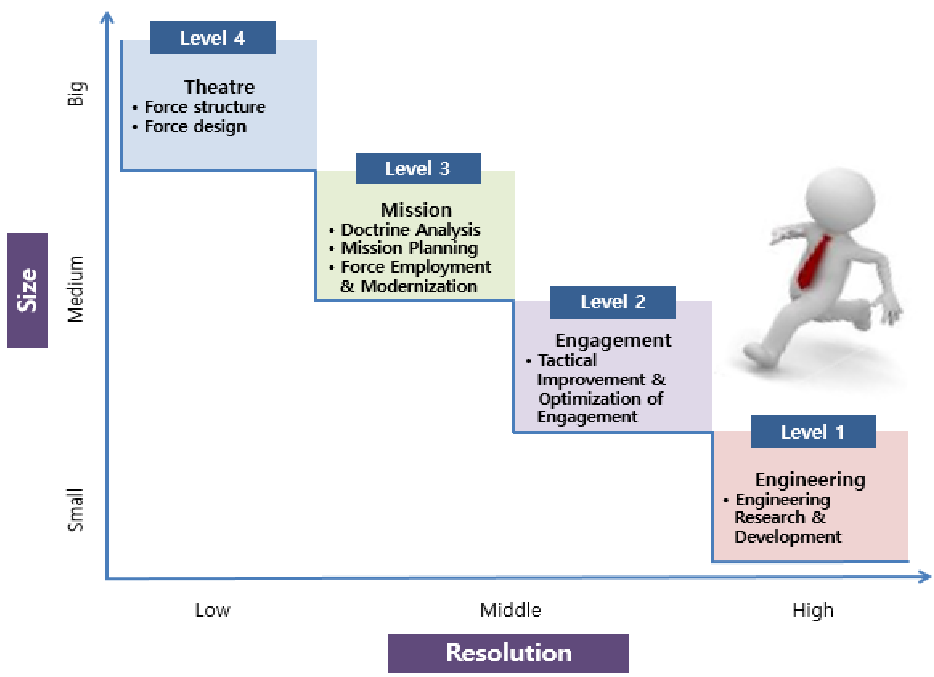

Figure 9. Four Principal Levels of Military Models

2.2.2.1 Engineering (Level 1)

The engineering model, which is physics-based and the closest to the real system in military models, is the technical level. Also, with the highest resolution the model simulates the entity level, such as an individual tank, airplane, radar, or its components. It is at this level that performance can be measured to design, test, and evaluate a variety of systems or subsystems in support of research and development of system through using engineering models. This model allows us to overcome the limitation of physical tests of a weapon system which is only executed under specific experiments with a few prototypes. This is because the engineering model enables us to take many experiments under the variety of scenarios with cost savings.

2.2.2.2 Engagement (Level 2)

This engagement model, in which several entities can be simultaneously simulated for the duel or the engagement, is the tactical level. The model simulates the regiment level consisting of individual entities with the high or middle resolution. It is used to design, test, evaluate, and improve various tactical operations so that the optimal operation can be chosen from the available operations.

2.2.2.3 Mission (Level 3)

This mission model, in which missions or battles can be simulated, is represented on the operational level. Corps or division level is simulated to analyze doctrine, design mission, or deploy and modernize forces through using the model, which is in the middle or low resolution.

2.2.2.4 Theatre (Level 4)

This theatre model is simulated on the strategic level, which is the highest level in military. The model simulates the national level with the low resolution, which shows one modeled unit including many entities. It is used to organize and design the force from the viewpoint of strategic level. This level supports the decision in the composition and the allocation of capabilities of forces.

2.2.3 Standard Simulation Architectures (SSAs)

The simulation models are usually standalone systems that are independently developed and operated. However, they can be integrated as a distributed system, which can be one of these types: live, virtual, and/or constructive simulation models, by SSAs using a network. This is because the integration of systems has a variety of benefits over standalone systems, such as higher performance, higher productivity, reusability, and reduced cost. SSA is known by some of the following: a) Distributed Simulation Architecture (Fujimoto, 1999; Loper & Cutts, 2010), b) Simulation Architecture (Gustavsson, Björkman, & Wemmergård, 2009), c) Distributed Simulation Protocol (Zalcman, Blacklock, Foster, & Lawrie, 2011), d) Modeling and Simulation Interoperability Standards (Tolk, 2012), e) Modeling and Simulation Interoperability Protocol (Granowetter, 2013), etc. The SSAs have been further enhanced since the late 1980's and are still being modified today through advanced main infrastructures for several DoD projects. The Simulation Networking (SIMNET) and the Aggregate Level Simulation Protocol (ALSP) programs, which were sequentially started in 1983 and 1989, were funded by Defense Advanced Research Projects Agency (DARPA) (Calvin et al., 1993; Weatherly et al., 1996). Although not being used anymore, the SIMNET was the basis for the Distributed Interactive Simulation (DIS) (Steinman & Hardy, 2004). In the mid-1990s the Defense Modeling and Simulation Office (DMSO) took on the High Level Architecture (HLA) project for combining advantages of ALSP and DIS into a SSA, which can support acquisition, analysis, and training fields (Steinman &

Hardy, 2004). In the early years of the 2000, the Test and Training Enabling Architecture (TENA) project was initially developed to incorporate live assets in the test range setting by the real-time test range community (Noseworthy, 2008). Also, the Common Training Instrumentation Architecture (CTIA) project was completed to assist the Live Training Transformation (LT2) product line of the U.S. Army (Lanman, Becker, & Samper, 2009). Usage percentage of each SSA as based on the Live Virtual Constructive Architecture Roadmap (LVCAR) survey is: a) SIMNET 0%, b) ALSP 5%, c) DIS and HLA 35%, d) TENA 15%, e) CTIA 3%, and f) others 7% (Gustavsson et al., 2009). Figure 10 illustrates the historical relationship, the comparison of characters, and the usage percentage in USA among main SSAs. Acronyms or abbreviations are the following: a) GOTS (GOvernment of The Shelf), b) IWG (Interface Working Group), c) API (Application Programmers Interface), d) COTS (Commercial Off-The-Shelf), e) IEEE (Institute of Electrical and Electronics Engineers), f) PDU (Protocol Data Unit), g) AMG (Architecture Management Group), h) SISO (Simulation Interoperability Standards), i) OMT (Object Model Template), j) AMT (Architecture Management Team), k) LROM (Logical Range Object Model), l) MW (Middleware), and m) Peo-Stri (Program Executive Office for Simulation, Training, and Instrumentation).

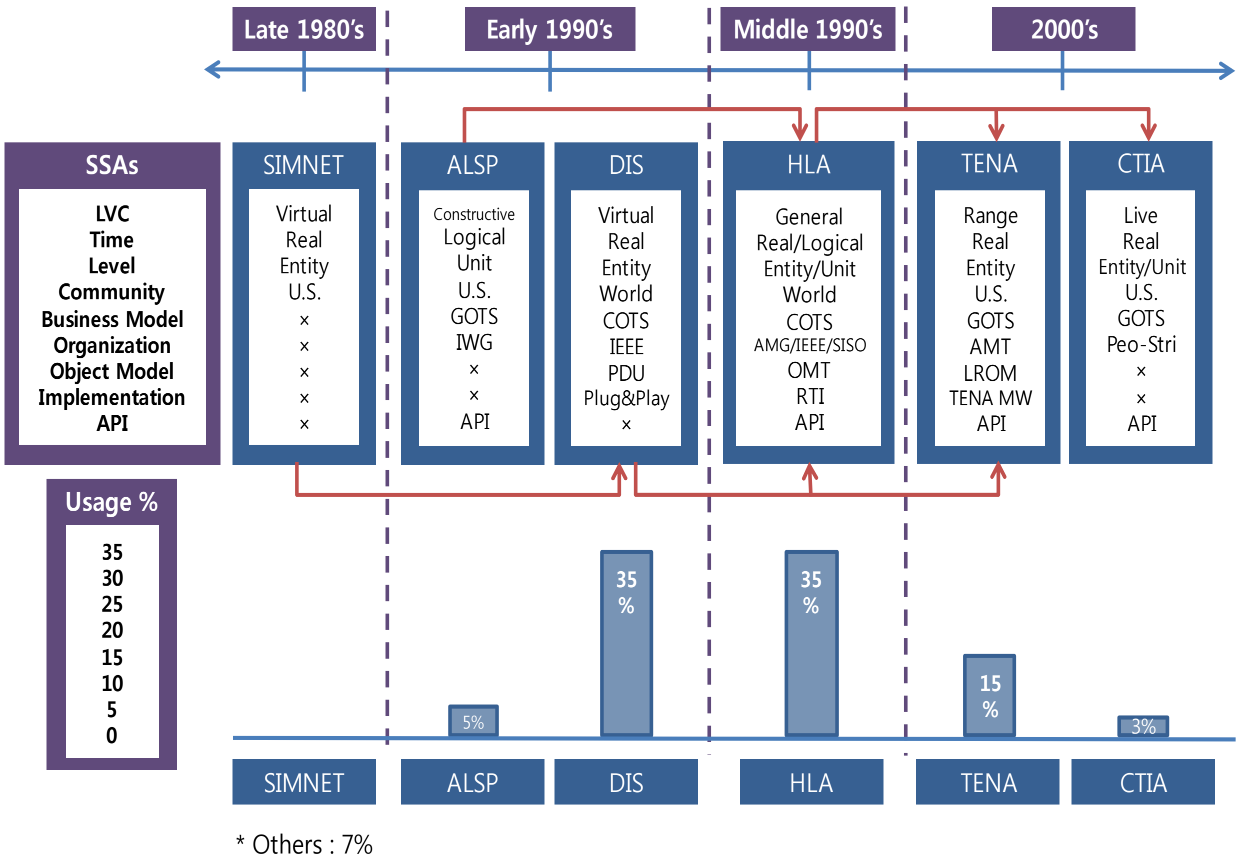

Figure 10. SSAs' Historical Relationship, Comparison of Characters, and Usage Percentage

2.2.3.1 Simulation networking (SIMNET)

This was the first successful architecture of SSAs for training in the military. It was developed to support the real-time distributed simulators, which are involved in the Combined Arms Tactical Training System (CATT), in order for the trainees to operate simulators simultaneously. Unlike the past, SIMNET enabled many simulators to engage together in the same virtual battlefield. The trainee operating a simulator receives and interprets the information regarding the other simulators' status in location and activity via a network, and then takes the appropriate action based on the analyzed results of communication among models (Calvin et al., 1993). The successful start of the SIMNET project led to the completion of the IDS project because of the incorporation of the SIMNET's main elements into DIS.

2.2.3.2 Aggregate level simulation protocol (ALSP)

The ALSP began from the need for the development of the new architecture to support synchronization among the aggregate level combat simulation models and to operate at a faster simulation time than at real world time for the theatre-level training (Weatherly et al., 1996). ALSP was successfully developed and, thus, a variety of war game models are presently operating in the army, navy, marines, and air force, and can be synchronized in order to execute a joint operation in an exercise.

2.2.3.3 Distributed interactive simulation (DIS)

The DIS's development was launched to support the virtual environment including Semi-Automated Forces (SAF) (Steinman & Hardy, 2004). The idea, which is to standardize the messages exchanged among simulators, gave rise to the IEEE 1278 DIS standard with the standardized PDUs. The PDUs are used for all communication of the entities and their interaction in the simulation (Tolk, 2012). From the viewpoint of a distributed system, DIS offers the plug and play function and does not need any other middleware and software. Therefore, DIS could be easily used on the spot. However, DIS is inefficient because PDUs are broadcasted to all simulation systems connected on the network even though they have nothing to do with some of the systems. Also, DIS requires entities to update the data every specific time interval even though their status does not change.

2.2.3.4 High level architecture (HLA)

The HLA was developed to enhance interoperability and reusability between the many various simulation systems executed in distributed environments (Steinman & Hardy, 2004). HLA consists of three main components, which are the framework and rules, the federate interface specification, and the object model template (IEEE, 2000). HLA is the distributed simulation architecture with the general purpose of a variety of systems and wide range of applications including analysis, training, acquisition, and logistic planning (Rabelo, Eskandari, Shaalan, & Helal, 2007). Also, HLA is an efficient SSA because individual simulation systems can filter whether the information is needed to be received or not for their systems at many various levels through the Runtime Interface (Hilbert, 2016; Tolk, 2012). Time management services are involved in HLA for ordering events as well as adjusting simulation time to fast or slow that are executed by means of synchronization algorithms (Tolk, 2012). However, in contradiction to DIS, HLA does not support the plug and play function and needs to take a middleware. Also, since some of the standards are left often in the Run-time Infrastructure (Hilbert, 2016) implementer, interoperability between different RTIs cannot be guaranteed (Blacklock & Zalcman, 2007). Moreover, a loosely connected federation could lead to tremendous cost for the simulation verification (Steinman, 2013).

2.2.3.5 Test and training enabling architecture (TENA)

TENA Software Development Activity (SDA) developed TENA to support the architecture and software implementation to satisfy the following: a) speedy and economical interoperability among facilities, range system, C4ISR (Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance) systems, and simulation systems, b) reuse for range asset utilization and for future developments, and c) composability to rapidly initialize, assemble, execute, and test a system from a pool of interoperable and reusable elements (Tolk, 2012). The TENA common infrastructure is a core part and includes the middleware, the repository, and the logical range data archive. The TENA object model specifies the interfaces and the common data shared by range resource applications. It includes many utilities, tools, and gateways to integrate faraway located range resources together in an appropriate manner (PEO-STRI, 2006). TENA has a significant benefit that is an auto-code generation that saves time required to integrate and test software in the range demonstration. Also, it has the enhanced capability to complete the routine tasks that are executed on the training and testing ranges in order to support exercises by using the auto-code generation, utilities, and common tools (Hudgins, 2007).

2.2.3.6 Common training instrumentation architecture (CTIA)

CTIA was developed to support a flexible product line architecture environment for the development and advancement of a common architecture in support of the U.S. Army's LT2 systems. It is composed of the standards, protocols, architecture services, and software components to be used by system developers. Also, it is the main software of the U.S. Army live Combat Training Instrumentation Systems (CTIS). It improves the training quality as well as reduces costs of development, logistics, maintenance, and training (Lanman et al., 2009).

2.2.4 Validation and Verification (V&V)

The validation and verification part explains the definitions and how to accomplish processes. The V&V methods are classified and then the specific methods are represented according to the classification.

2.2.4.1 Definition and process

The V&V of a simulation model is a very important part because the validity of the simulation results cannot be accepted without V&V. Verification is the process of certifying that the conceptual model has been implemented into a computer model with satisfactory accuracy (Davis, 1992) in building the model correctly. On the other hand, validation is the process of certifying that the model has sufficient accuracy for the goal (Carson, 1986); namely, building the right model. Sargent studied one of the most referenced V&V processes (Sargent, 2011). First, the conceptual model validation is executed repeatedly until satisfaction and then the conceptual model is built. Second, the computerized model verification is executed repeatedly until satisfaction and then the computerized model is built. Third, operational validity is executed by comparing the computerized model with the problem entity. If the requirements of the model changes are identified, the conceptual or computerized model must be executed again until verification and validation are satisfactory. Figure 11 summarizes the above V&V processes: a) the problem entity is the real or proposed system, situation, idea, phenomena or policy to be modeled, b) the conceptual model is the mathematical, logical, and verbal representation of the problem entity developed for a specific study through an analysis and modeling step, and c) the computerized model is the conceptual model developed on a computer through a computer programming and implementation step.

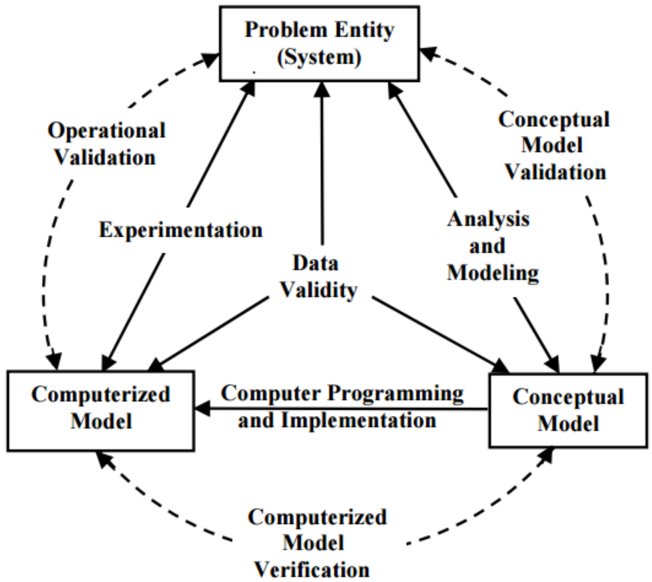

Figure 11. Simplified Version of the Modeling Process (Sargent, 2011)

2.2.4.2 Methods

These are the methods categorized to support V&V as the following: a) informal testing methods are more subjective approaches to be assessed using human intuition and evaluation without mathematical formalism and can be used as a reference, b) static testing methods are non-dynamic assessment approaches using characteristics of the model design and code without simulation execution, c) dynamic testing methods are dynamic approaches to be assessed using the simulation execution and evaluation of its results compared to other observations and models in the experiments executed in the real world, and d) formal testing methods are more objective approaches to be assessed using mathematical proofs. Figure 12 represents V&V methods that Balci categorized (Balci, 2007).

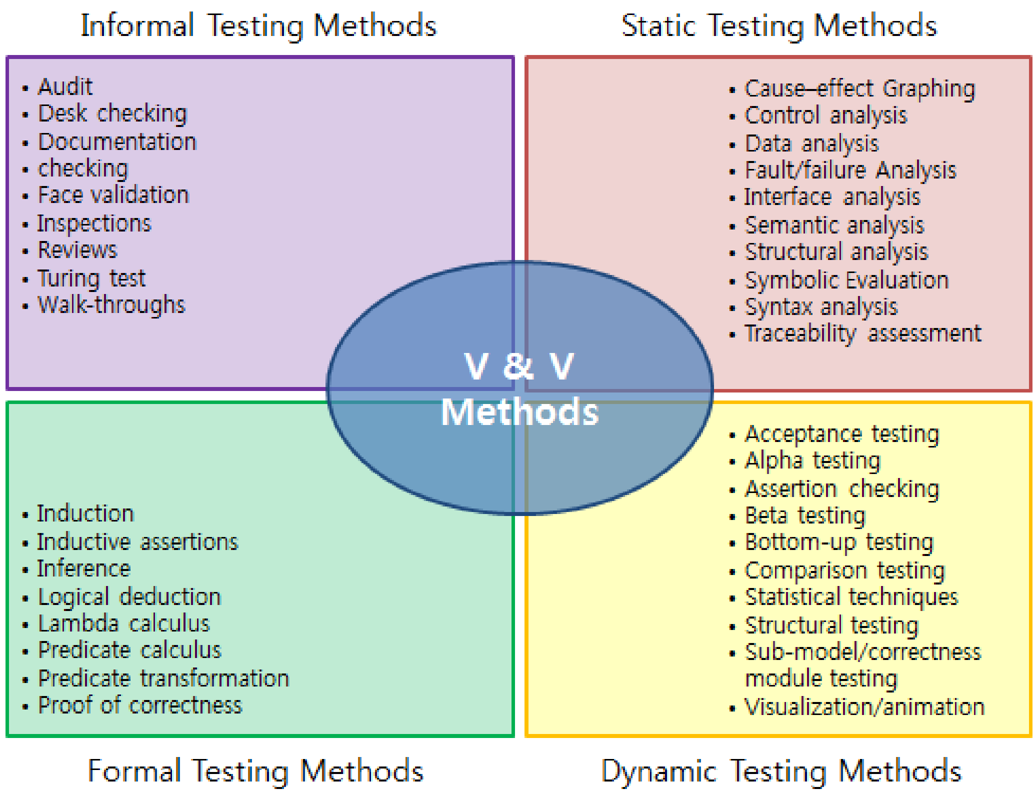

Figure 12. V&V Methods

2.3 Big Data

The 'Big Data' concept emerged not from the academic field but from other unknown sources. Diebold (2012) claims that "Big Data . . . probably originated in lunch-table conversations at Silicon Graphics Inc. (SGI) in the mid-1990s, in which John Mashey figured prominently". The term 'Big Data' was primarily used as a keyword in "Big data: The next frontier for innovation, competition, and productivity" and the term has not only represented the pronoun but also became the trend of the modern age (Manyika et al., 2011). In 1970 the term 'Big Data' was referred to the development of an algorithm for processing big data. Since then the concept has changed. The big data research began in earnest in 2012 as shown in Figure 13. The current trend was caused by the promotional plan from IBM and other technology companies, such as google, amazon, and e-bay, which invested in forming the niche analysis market. In the following section, we review general big data concepts, characteristics, and analysis techniques and technologies studied until the present. The DBD are further reviewed in detail.

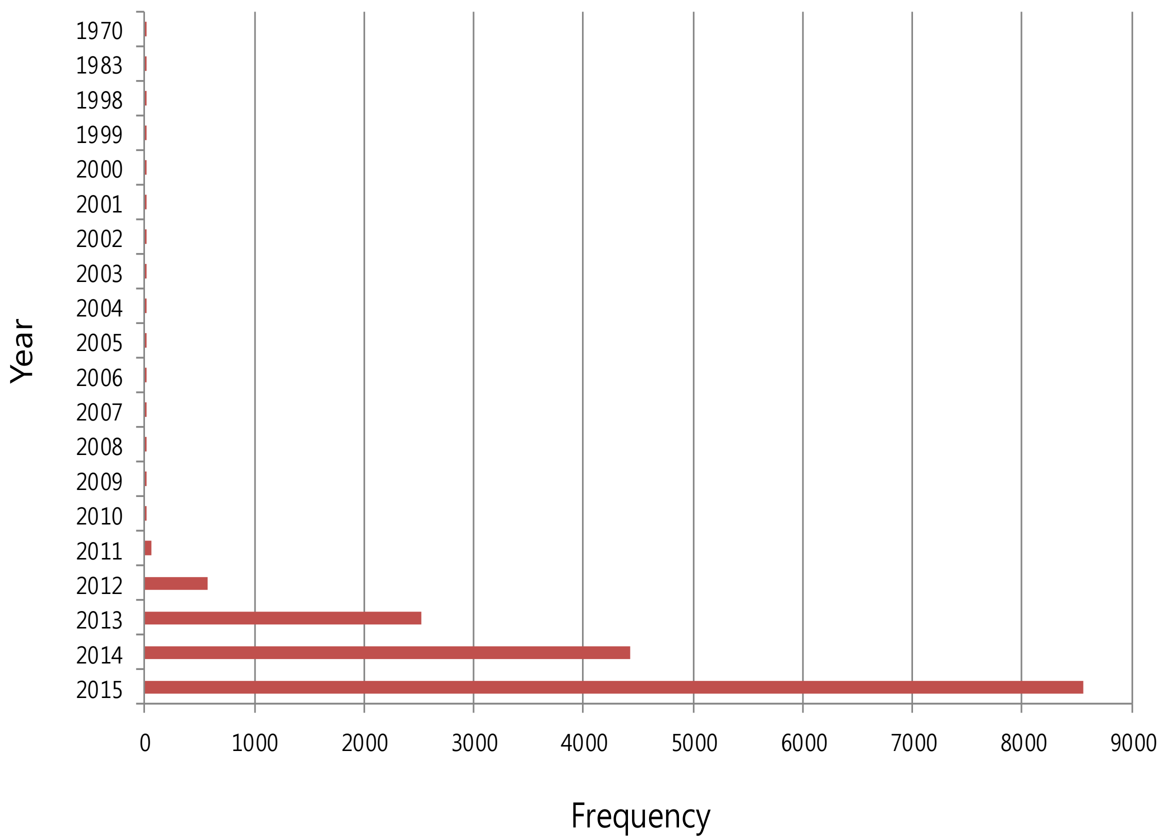

Figure 13. Frequency Distribution of Research Including Term "Big Data" based on Compendex Database

2.3.1 General Big Data

2.3.1.1 Concepts

Since the big data concepts have emerged, like many others with commercial purposes, without thorough academic research, there are a variety of definitions which apply according to Gandomi and Haider (2015). They reviewed the big data definition based on an online survey of 154 global executives and concluded that the main definitions are the following: a) "new technologies designed to address the volume, variety, and velocity challenges of big data", b) "massive growth of transaction data, including data from customers and the supply chain", c) "requirement to store and archive data for regulatory and compliance", and d) "explosion of new data sources, such as social media, mobile device, and machine-generated devices" (Gandomi & Haider, 2015).

2.3.1.2 Characteristics

Laney (2001) suggested that big data is characterized by volume, variety, and velocity (The Three Vs). Also, the three Vs are the core characters of a common framework in support of representing big data (Hsinchun, Chiang, & Storey, 2012). International Data Corporation (IDC) suggested that big data have four characteristics: volume, variety, value and velocity (The Four Vs) (Meng, 2012). Hilbert (2016) suggested five Vs: volume, variety, velocity, variability, and veracity (The Five Vs) as characteristics of big data. Big data can be characterized by the six Vs based on the above research: a) Volume refers to the huge magnitude of data more than 1 TB, b) Variety means the heterogeneity of data types, such as structured, unstructured, and semi-structured data, c) Velocity presents high data generation and processing speed, d) Variability means complexity by the variation in the data occurrence rates because big data are generated by a variety of sources, e) Value refers to the low value density compared to amount of data, and f) Veracity means unreliability inherent to subjective data such as customer private taste. Since these characteristics can be affected by technical progress and applied field and environment, they are not fixed but evolve continuously.

2.3.1.3 Analysis techniques and technologies

Big data analysis techniques have a few differences within the general data analysis. The big difference between them is that special technologies are required to aggregate, store, and process big data for overcoming its limitations because the problems cannot be solved based on the characteristics only with past computing technologies. For example, if the data are huge, unstructured, and an high speed process is required, current technologies can be used in order to overcome analysis limitations with the capacity of the general computer processing, such as Hadoop, NoSQL, and stream processing. McKinsey Global Institute introduced a variety of techniques and technologies for big data analysis as shown in Table 3 (Manyika et al., 2011) .

Table 3. Big Data Analysis Techniques and Technologies

Techniques (26)		Technologies (24)	
A/B testing	Association rule learning	Big Table	Business intelligence
Classification	Cluster analysis	Cassandra	Cloud computing
Crowdsourcing	Time series analysis	Data mart	Data warehouse
Data mining	Ensemble learning	Dynamo	Distributed system
Optimization	Machine learning	Hadoop	Extract, transform, and load
Neural networks	Natural language processing	HBase	Google File System
Network analysis	Genetic algorithms	MapReduce	Mashup
Spatial analysis	Predictive modeling	Metadata	Non-relational database
Regression	Sentiment analysis	R	Relational database
Signal processing	Pattern recognition	SQL	Semi-structured data
Statistics	Supervised learning	Stream processing	Structured data
Simulation	Unsupervised learning	Unstructured data	Visualization
Visualization	Data fusion and data integration		

2.3.2 Defense Big Data

The United States DoD established special mission organizations to systemically collect, manage, and analyze data. Representatives of these are the Center for Army Lessons Learned (CALL) and Army Materiel Systems Analysis Activity (AMSAA). CALL continuously conducts the Army Lessons Learned Program. Also, it identifies, assembles, analyzes, distributes, and documents lessons and best practices while sustaining global situational recognition in support of sharing knowledge and facilitating the Army's and Unified Action Partners (UAPs)' adaptation in order to succeed in military confrontations. As the multi-media based operation, CALL distributes lessons and related materials via various electronic and print media. AMSAA's net goal is to offer the military the best U.S Army materials. AMSAA aids the U.S. Army via systems and engineering analyses in a support of making decisions for material's acquisitions, technology, and the planning, developing and maintaining of the U.S. Army weapon systems. Particularly, Joint Technical Coordinating Group for Munitions Effectiveness (JTCG/ME) produces Joint Munitions Effectiveness Manuals (JMEMs) that offer tri-service effectiveness data for non-nuclear U.S. weapons at the engineering level.

There are a few cases generating big data using DM&S for training and operations analysis. For example, STOW97 was developed as a distributed simulation consisting of hundreds of computers for military training and generated 1 TB data during 96 hours (Wilcox, Burger, & Hoare, 2000). Joint Semi-Automated Forces (JSAF) was studied for the joint urban operation. Hundreds of nodes for simulating JSAF were connected with geographical sites, and 3.7 TB data was generated by the model (Graebener, Rafuse, Miller, & Yao, 2004). The U.S. Naval and Air Force use the Synthetic Theater Operations Research Model (STORM), which simulates the campaign level to effectively plan the mission by finding the main factors. STORM can create gigabyte data from one replication and then larger data can be generated according to the scenario and model complexity (McDonald et al., 2014). However, the above cases were not analyzed entirely by using the concept of big data but used a portion of big data without its techniques.

2.4 Gap Analysis

This section compares the proposed methodology with existing research by analyzing gaps between them. Proper criteria for the gap analysis are suggested and then existing research is reviewed based on the criteria.

2.4.1 Criteria

This part extracts the standards by which gaps between the proposed methodology and existing research are analyzed. Specific elements that have been included and made for improving the fidelity of WCE analytics are identified as the criteria of gap analysis. Also, other elements, which should be considered to overcome the weaknesses of the existing research related to WCE analytics using modeling and simulation techniques, are chosen for the comparison criteria between the proposed and existing research.

Although existing research has tried to enhance the fidelity of WCE analytics, it has not considered the utility of abundant data and improvement of modeling reality by which the fidelity of WCE analytics can be boosted. From the viewpoint of the utility of abundant data, the experiments for the WCE analytics have found many limitations because of the lack of times it has been able to be applied in the real world to obtain abundant data. Currently, the utility of LVC simulations can be a solution to obtain abundant data. Also, the data can be analyzed through the utility of big data techniques. From the viewpoint of improvement of modeling reality, most research has paid little attention to minimizing assumptions, considering various factors, and applying a variety of scenarios. They have had limitations for maximizing the fidelity of WCE analytics without the above three elements. Therefore, the limitations can be solved through focusing on the elements in advance. Existing research also has other weaknesses except for the above elements of WCE analytics. Most methodologies for the WCE analytics are not only for a specific situation but for limited analytics. Since the methodologies are not generalized, the limitation of their application can exist. Also, the relationship only between a single factor and MOE is analyzed in most of the existing research, and thus, the comprehensive analysis cannot be executed. Analytics results are not sufficient to support decision-making. This is because only partial analytics results are drawn without considering various factors simultaneously which can be presented by the WCE equation. Also, most of the existing research recommends optimal values for WCE without constraints; therefore, the analytics results are incomplete. Moreover, the optimal values recommendation is needed based on WCE equation and the constraints to enhance the completeness of analytics results.

Therefore, based on the above rationale, criteria of gap analysis are identified as shown in the following: a) utility of abundant data; a-1) utility of LVC simulations benefits, a-2) utility of big data technique and technology benefits, and a-3) external source utility, b) modeling reality; b-1) minimizing assumption degree, b-2) various factors application, and b-3) scenario variety, c) generalization; c-1) flexibility of application, and c-2) comprehensive analysis, d) analytics result usability; d-1) WCE equation estimation with various factors, and d-2) optimal values recommendation based on the constraints.

2.4.2 Analysis of Existing Research

This part reviews the existing research related to WCE analytics only using the M&S technique in order to identify practical research gaps because there has been no existing research published using big data for WCE analytics until recently.

Connors (2015) has suggested a methodology, which consists of seven steps: a) define metrics, b) modeling environment, c) scenario description and simulation development, d) Advanced Framework for Simulation, Integration and Modeling (AFSIM) scenario implementation, e) agent decision making behavior, f) verification and validation, and g) experimental design and analysis. The goal of this research is to obtain potential benefits for a new weapon system by analyzing the main parameters influencing the weapon effectiveness. As a case study, the Cuda prototype, which had been designed by the Lockheed Martin Corporation, for the substitute of the Small Advanced Capability Missile (SACM) was used to apply the methodology. This study suggested various MOE examples for WCE analytics and extracted the main factors influencing each MOE. However, only two tactics were considered for WCE analytics for the new weapon based on the anticipated main factors, which were not identified within a variety of scenarios but only assumed by author. Also, this paper's methodology was limited to acquisition of the new weapon system within the Air Force. The optimal values were not recommended for acquisition of the new weapon as this study has only analyzed the effectiveness according to whether the new weapon is loaded or not.

J. Lee, Shin, Kim, Bae, and Kim (2015) studied the combat effectiveness of an Unmanned Ground Vehicle (UGV) using ABS. This paper defined a MOE as the blue team's survival ratio and MOPs as detection and hitting ratio. To identify interrelation between a MOE and MOPs, ABS was built with three types of weapon systems, which are the UGV, the tank, and the armored vehicle, and simulated the model according to various MOPs' levels under a given scenario. A sensitivity analysis was done to compare the MOE with MOPs. This paper's contribution is to suggest a WCE equation, which proves that the detection rate is a more important factor in increasing the blue team's survival rate than the hitting rate. However, this study only considered two MOPs to measure the effectiveness of the UGV under one scenario. The result has limitations to represent the effectiveness of the real systems. Furthermore, although the detecting and hitting rate ranges are limited due to technical and budget problems, this study did not reflect the limitation, and thus, optimal values were not identified. Also, this paper did not consider various data to analyze the effectiveness of UGV and suggest the method for applying generalized cases.

Jiao, Li, Ma, and Yang (2013) suggested the simulation evaluation system for weapon operational effectiveness. The authors presented three steps for the evaluation of weapon operational effectiveness such as following: a) extracting the data for depicting the system character from simulation experiment data via relation operation, b) gaining the value of evaluation index via arithmetic operation, functional operation, or other complex operations, and c) gaining the operational effectiveness via integrating the evaluation indexes. This process can be adapted to a variety of fields. The study focused on the knowledge management for the weapon effectiveness, of which the analytics method was not shown. Therefore, there are a variety of limitations for the realistic WCE analytics with this research.

Yoo, Lee, and Baik (2012) have built a methodology for future weapon system effectiveness analytics using the One Semi-Automated Forces (OneSAF) based on the Product Line Architecture Framework (PLAF). The methodology processes consist of the following six steps: a) characteristics analytics of future weapon systems, b) modeling future weapon systems, c) model implementation of future weapon systems, d) composition of future weapon systems, e) simulation, and f) result analytics. Their study explained that OneSAF, which consists of various hierarchical function components, is an appropriate model to analyze future weapon system effectiveness. Through the case study, the researchers proved that the methodology using a PLAF based simulation system is better than the conventional one using the Janus model because of showing more realistic results. Furthermore, this paper focuses on six scenarios to improve the fidelity. However, this study did not analyze the broad aspects because it did not consider all factors influencing WCE. Moreover, the limited factors were not reflected for comprehensive analytics, and the optimal values of factors to maximize WCE were not suggested in this paper.

Seo et al. (2011) have evaluated the WCE on the anti-torpedo combat systems through the constructive simulation, which is an underwater warfare simulation based on the hierarchical formalism that is a DEVS. This study includes four kinds of weapon systems which are submarine, torpedo, surface ship, and decoy. Each weapon system has one coupled model which is an entire platform model. It includes three atomic models that represent the main functions of the entire model. Also, each atomic model consists of two layers as shown in the following: a) Discrete Event System (De Maesschalck, Jouan-Rimbaud, & Massart) layer which describes the abstract process of an atomic model, and b) Object Model (OM) layer which details process of an above atomic. The model was implemented by C^{++} with $DEVSim^{++}$ library. This study considered several factors, which are the decoy operating system patterns, the detection range, the decoy operating time, and the mobile decoy speed, and investigated their influences to MOEs. This study analyzed MOEs under only one scenario with a variety of assumptions based on the limited information, and thus, the results have a restriction in terms of generalizing analytic results of MOEs. Also, the method shows a process on a specific case. This study extracts the optimal values of factors influencing WCE without constraints. Therefore, the values are suspicious on whether they are real or not.

C. Jung and Lee (2010) suggested the methodology on the effectiveness based analysis for the optimal requirement of the new weapon using ABS. This study shows how to estimate the optimal requirement of the new attack helicopter against anti-armored corps using Army Aviation simulation (AAsim). The methodology for the attack helicopter effectiveness analytics follows the following six steps: a) identify and input data for simulation, b) input scenario, c) simulation, d) result analytics, e) calculate the combat effectiveness, and f) estimate the optimal requirement. This paper proposed the general methodology for the effectiveness analytics of the new weapon and number of optimal requirements for the attack helicopter according to the number of anti-armored corps. However, the research did not represent the WCE equation. Therefore, the optimal values cannot be extracted under various conditions. Also, the authors did not consider various data, which can be obtained from other sources, and thus, missed the opportunity to improve the accuracy of the results.

K. Liang and Wang (2006) analyzed the effectiveness of mix strategies of decoy and jammers in anti-torpedo tactics using simulation and evolutionary algorithms. This paper built the simulation model to evaluate the tactics by analyzing interaction among the torpedo, the jammers, the submarine, and the decoy. The model was simulated based on the instruction of the predetermined tactics. At this time, the MOE was decided to be the success rate of the submarine evasion. This research used the evolutionary algorithms to optimize the tactics. It is important to suggest the method to optimize tactics. However, this paper considered scenarios for only four directions of the torpedo attacks to analyze the effectiveness of the tactics so that limited solutions were suggested. Also, this research did not use various data for the effectiveness analytics. Furthermore, the method was fixed only for analyzing the effectiveness of tactics according to the scenarios.

C. M. Anderson (2004) has suggested “generalized” weapon effectiveness modeling. This paper showed the weapons effectiveness for several cases by contrasting the JMEM to Monte Carlo simulation, which applied Carleton Damage Function or Rectangular Cookie Cutter Approximation. Anderson analyzed weapon effectiveness using JMEM and Monte Carlo simulation for the following scenarios: a) Single vs. Unitary Target, b) Single Weapon vs. Area Target, and c) Stick of Weapon vs. Area Target. Chusilp et al. (2014) analyzed the weapon effectiveness using the Pk matrix and Carleton damage function with Monte Carlo Simulations. This paper measured the probability of damage by the Pk matrix and Carleton damage function according to the type of targets, the number of artillery weapon shots, and the number of simulation runs. This research represented how to analyze the weapon effectiveness in engineering level, and thus, the result can be used to build the various simulation models of engagement level. However, since the WCE analytics at the engineering level did not consider various factors and scenarios, the model reality was insufficient. Also, these studies did not use a variety of data to improve the fidelity of results. Furthermore, the WCE equations and the optimal values were not suggested for the decision maker to determine alternatives effectively and efficiently.

Armo (2000) represented the relationship between a submarine's maximum speed and its evasive capability using the discrete event simulation. This paper simulated three systems, which were a submarine, a countermeasure system, and a light Anti-Submarine Warfare (ASW) torpedo. As the MOE for the submarine was the probability of not being caught by the ASW torpedo, it was calculated with changing the maximum speeds according to the attacks of the ASW torpedo. This research contributed to identifying that the maximum speed of a submarine relates to its evasive probability within a specific range of speed using the discrete event simulation. However, since this study considered only the maximum speed under a specific scenario to measure MOE, many limitations are followed in the application to the results in the real system. Also, the specific method for the research was used, and thus, the method has a limitation for analyzing the effectiveness of various weapons systems.

2.4.3 Research Gaps

This part summarizes research gaps between the existing research and this dissertation in terms of WCE analytics based on the comparison criteria. The existing research has limitations such as the following: a) overlook the benefit of utility of abundant data, b) lack the ability to build a simulation model similar to the real combat system, c) unable to generalize methods for a variety of fields related to WCE analytics, and d) inability to represent the completeness of analytics results for decision makers to use effectively and efficiently. This research fills in the previous research gaps as shown in Table 4.

Table 4. Summary of Research Gap

Existing Research		Comparison Criteria									
		Utility of Abundant Data			Modeling Reality			Generalization		Completeness of Analytics Results	
		Utility of LVC simulations benefits	Utility of Big Data Technique and Technology benefits	External Source Utility	Minimizing Assumption Degree	Various Factors Application	Scenario Variety	Flexibility of Application	Comprehensive Analysis	WCE Equation Estimation with Various Factors	Optimal Values Recommendation
K. R. Armo (2000)		N	N	N	W	W	W	W	W	N	W
C. Anderson (2004)		N	N	N	N	N	M	M	W	N	W
Liang et al. (2006)		N	N	N	W	W	W	W	W	N	M
C. Jung et al. (2010)		N	N	N	W	M	W	M	W	N	M
K. Seo et al. (2011)		N	N	N	W	M	W	W	W	N	W
S. Yoo et al. (2012)		N	N	N	W	W	M	M	W	N	N
S. Jiao et al. (2013)		N	N	N	N	N	N	M	W	N	N
P. Chusilp (2014)		N	N	N	N	N	N	M	W	N	W
J. Lee et al. (2015)		N	N	N	W	W	N	W	W	M	N
C. Connors (2015)		N	N	W	W	W	W	M	W	N	W
Jung (2018)	Method -ology	M	M	M	M	S	S	S	S	S	S
	Case Study	W	M	W	M	M	M	S	S	S	S

Not Covered Weakly Covered Moderately Covered Strongly Covered

CHAPTER THREE: METHODOLOGY

This chapter explains the new methodology for the WCE analytics using big data generated by LVC simulations. The procedure is suggested for the WCE analytics methodology. The WCE analytics system diagram and formalism are proposed to give shape to the procedure. Also, the procedure is explained step by step in each section in detail.

3.1 Overview

Until present time, weapon effectiveness analytics has been done based on limited environments which emphasized: basic constructive simulations, one weapon system, engineering level, and limited information. A more realistic study must analyze the level of complexity of the weapon system and the numerous variables under joint battlefield environments. Of course, if the WCE is analyzed using the abundant data collected from experiments or results of warfare done in the real world, it is not only the best realistic but also the most reliable study. However, it is not possible because warfare data for analyzing WCE are insufficient as well as WCE analytics from experiments in real world are a very costly, time-consuming, and dangerous tasks. Also, the experiments could lead to diplomatic conflicts for nearby countries. To overcome the above mentioned problems on the analysis of complex weapon systems, it is a given that the new methodology on WCE analytics will use the big data generated by LVC. This is because LVC simulations have the potential to provide more solutions to the structure of the problem by reflecting a more realistic environment.

There are two ways to use big data for WCE analytics. They include the following: a) utilizing big data generated by new simulation models and b) utilizing big data accumulated by existing simulation models. The unified common analytic platform is needed to analyze big data generated by the above two ways as shown in Figure 14. The second way can reduce the model development time; whereas the first way enables us to generate the intended factors for analyzing WCE.

Figure 14. Concept for Unified Platform to Analyze Big Data Generated by Simulation Model

From the viewpoint of the first one, the framework for a new methodology is suggested to overcome the above problems and to enhance the “fidelity” for the weapon effectiveness analytics using the big data generated by Six Degrees of Freedom (6DoF) LVC simulations as shown in Figure 15. This framework consists of five steps, which are generating big data (step 1), collecting additional data related to WCE (step 2), processing big data for WCE analytics (step

3), estimating the WCE equation and optimal values (step 4), and reporting results (step 5). All steps proceed in regular sequence, but each step can be returned to the previous one if necessary. At this moment, steps 1, 2, and 3 are the data pool preparation level to generate, collect, and process big data for performing analytics. Step 4 the analytics level plays a role as the unified data analytics platform based on the big data generated, collected, and processed as mentioned above. Step 5 the visualization level enables decision makers to efficiently and effectively select an optimal alternative under limited conditions.

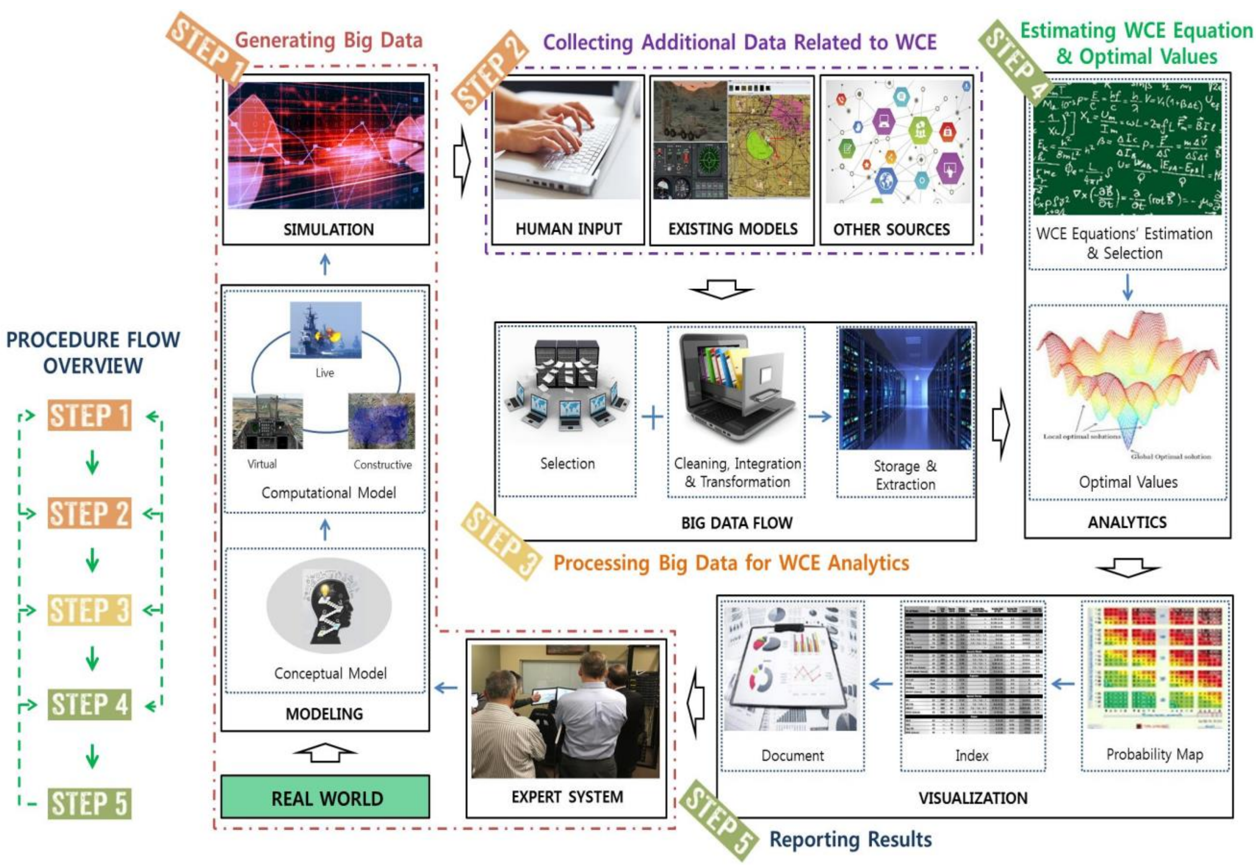

Figure 15. Framework on WCE Analytics using Big Data Generated by LVC Simulations

3.2 WCE Analytics System Diagram and Formalism

The developed diagram is a specific modeling language for the WCE analytics system. It specifies the WCE analytics system using a sequential, hierarchical, behavioral, relational, and modular form. It is intended to support the specification of this complex system for stakeholders to easily understand. Furthermore, this diagram displays a useful modeling language which can be applied to modification, extension, simplification, new design, verification, validation, and accreditation of the WCE analytics system. Figure 16 and Figure 17 show the WCE analytics system diagram.

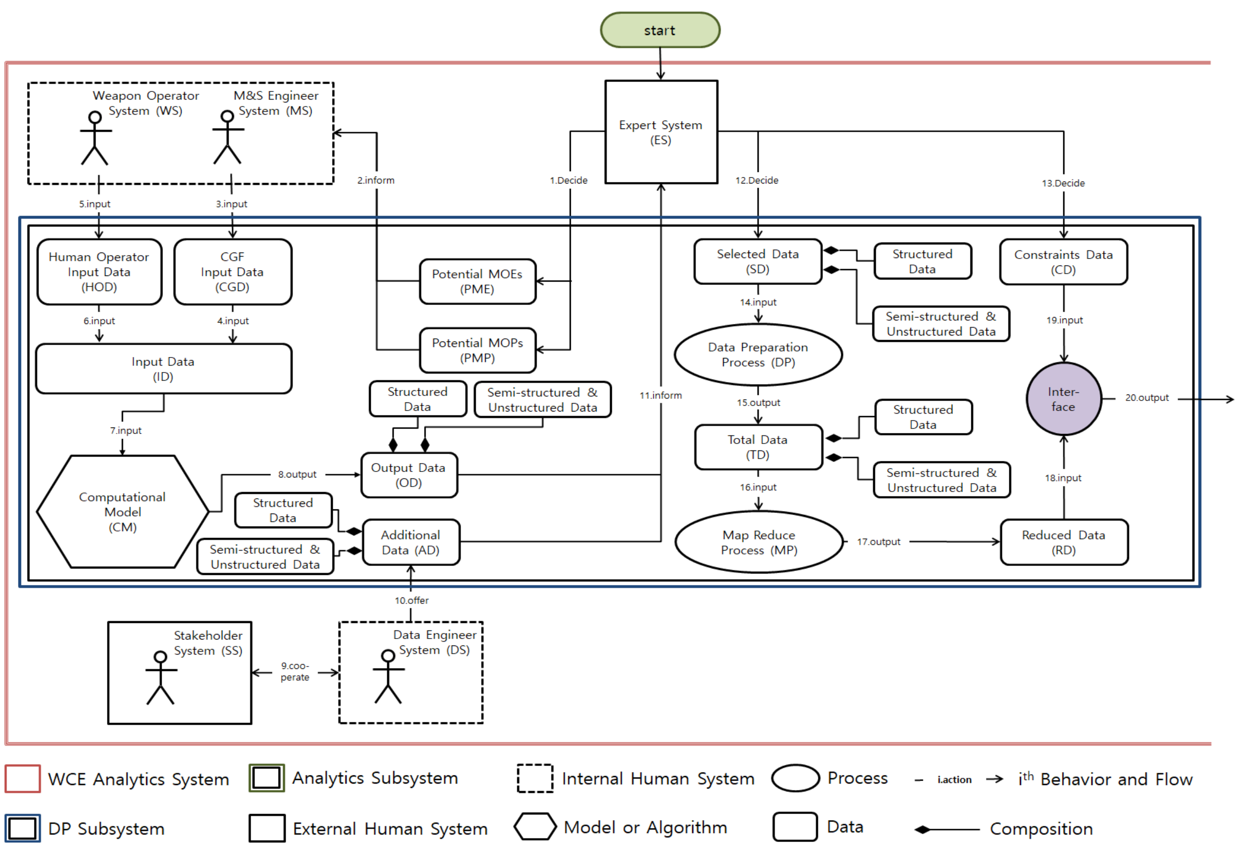

Figure 16. WCE Analytics System Diagram including Data Preparation Subsystem

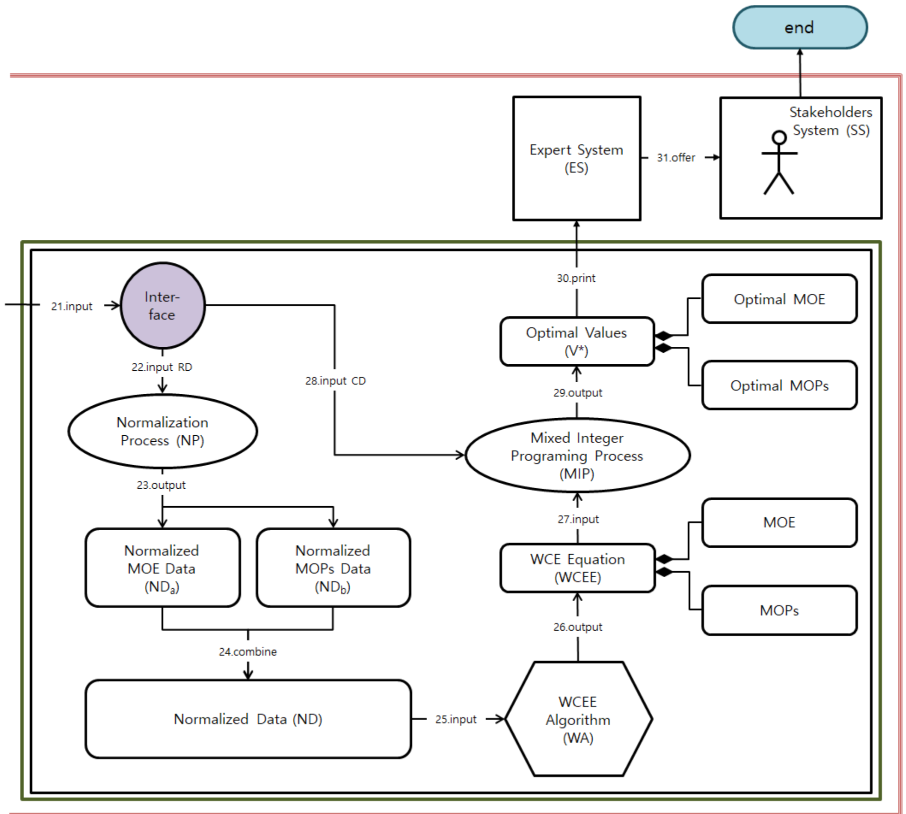

Figure 17. WCE Analytics System Diagram including Analytics Subsystem

The WCE analytics system has two subsystems and five other systems which are the data preparation and the analytics subsystems as well as an expert system, a weapon operator system, an M&S engineer system, a data engineer system, and a stakeholder system. The WCE analytics system related to the data preparation subsystem has the following flows: a) The expert system decides the potential MOEs and the potential MOPs and then sends the information to the weapon operator system and the M&S engineer system, b) The M&S engineer system inputs data for modeling Computer Generated Force (CGF) to the computational model, c) The weapon operator system inputs the human operators' data for operating the simulator to the computational model, d) The output data is generated by the computational model, e) The data engineer system collects two types of additional data which are for MOPs and constraints under the cooperation of the stakeholder system and then offers the additional data, f) The output data and the additional data is sent to the expert system, g) The expert system selects data to be used for the WCEE development, h) It decides the constraints data which is used as conditions to find optimal values, i) The selected data are prepared for the total data, which is stored to HDFS, by the data preparation process, j) The reduced data is extracted from the total data by the MapReduce process, k) The reduced data and the constraints data are gathered to the interface of the data preparation subsystem, and l) The gathered data comes out to the interface of the analytics subsystem. The WCE analytics system associated with the analytics subsystem has the following flows: a) The reduced data and the constraints data are transferred to the interface of the analytics subsystem, b) The reduced data is processed to the normalized MOEs data and the normalized MOPs data by the normalized process, c) The normalized MOEs data and the normalized MOPs data are combined to the normalized data, d) The normalized data is used as the input data of WCEE algorithm, e) The WCEE, which is composed of the MOEs and the MOPs, is developed by the WCEE algorithm, f) The WCEE and the constraints data are used as inputs of the mixed integer programing process, g) The optimal values, which have the optimal MOEs and MOPs, are estimated by the mixed integer programing process, h) The expert system confirms the optimal values, i) Then it gives them to the stakeholder system, and j) The stakeholder system has the final decision on the results.

The WCE analytics formalism is a set-theoretic formulation which is a set-mathematical logic. The formalism specifies the WCE analytics system using a sequential, hierarchical, logical, mathematical, relational, and modular form. It is intended to support the diagram by offering a logical and mathematical form which does not belong to the diagram's characteristics. The formalism has three models which are an Upper Level Model (ULM) for the WCE analytics system as well as a type of Lower Level Model (LLM) for the data preparation subsystem, and another type of LLM for the analytics subsystem. The ULM involves two LLMs and supports the LLMs to completion of their functions while operating. The ULM is specified by a 13-tuple as shown in Equation (3.1).

$$\mathrm{ULM} = < \theta_{ES}, LLM_{DP}, \xi_{MS}, \xi_{WS}, LLM_{DP}, \theta_{SS}, \xi_{DS}, LLM_{DP}, \theta_{ES}, LLM_{DP}, LLM_{AN}, \theta_{ES}, \theta_{SS} > \tag{3.1}$$

θ_{ES}: Expert System

LLM_{DP}: LLM for the Data Preparation Subsystem

ξ_{MS}: M&S Engineer System

ξ_{WS}: Weapon Operator System

θ_{SS}: Stakeholder System

ξ_{DS}: Data Engineer System

LLM_{AN}: LLM for the Analytics Subsystem

The LLM_{DP} explains the subsystem which has a function for the data preparation. It has nine data sets, one model, and three processes to achieve its function. Furthermore, it receives support from five systems which are the expert system, the M&S engineer system, the weapon operator system, the stakeholder system, and the data engineer system. The LLM_{DP} is specified by a 13-tuple as shown in Equation (3.2).

$$\mathrm{LLM_{DP}} = < PME, PMP, ID, \delta_{CM}, OD, AD_a, AD_b, SD, CD, \psi_{DP}, TD, \psi_{MP}, RD > \qquad (3.2)$$

PME: Potential MOEs Set ($\theta_{ES} \rightarrow PME$)

PMP: Potential MOPs Set ($\theta_{ES} \rightarrow PMP$)

ID: Input Data Set (ξ_{MS}, ξ_{WS}: $PME, PMP \rightarrow ID$)

δ_{CM}: Computational Model

OD: Output Data Set (δ_{CM}: $ID \rightarrow OD$)

AD_a: Additional Data Set for MOPs ($\theta_{SS}, \xi_{DS} \rightarrow AD_a$)

AD_b: Additional Data Set for Constraints ($\theta_{SS}, \xi_{DS} \rightarrow AD_a$)

SD: Selected Data Set (θ_{ES}: $OD, AD \rightarrow SD$)

CD: Constraints Data Set (θ_{ES}: $AD \rightarrow CD$)

ψ_{DP}: Data Preparation Process

TD: Total Data Set (ψ_{DP}: $SD \rightarrow TD$)

ψ_{MP}: MapReduce Process

RD: Reduced Data Set (ψ_{MP}: $TD \rightarrow RD$)

The data analytics subsystem is described by the $\mathrm{LLM}_{\mathrm{AN}}$ which has three data sets, one algorithm, and two processes to complete the data analytics to be offered to the stakeholder system. Also, two systems, which are the expert system and the stakeholder system, support the $\mathrm{LLM}_{\mathrm{AN}}$ by checking the data analytics result if necessary. The $\mathrm{LLM}_{\mathrm{AN}}$ is specified by a 6-tuple as shown in Equation (3.3).

$$\mathrm{LLM}_{\mathrm{AN}} = < \psi_{NP}, ND, \lambda_{WA}, WCEE, \psi_{MIP}, V^{*} > \tag{3.3}$$

ψ_{NP}: Normalization Process

ND: Normalized Data (ψ_{NP}: $RD \rightarrow ND$)

λ_{WA}: WCEE Development Algorithm

$WCEE$: Weapon Combat Effectiveness Equation (λ_{WA}: $ND \rightarrow WCEE$)

ψ_{MIP}: Mixed Integer Programing Process

V^{*}: Optimal Values' Set (ψ_{MIP}: $WCEE, CD \rightarrow V^{*}$)

This formalism based on the WCE analytics diagram is applied to supporting specification on the new complex WCE analytics methodology as the specific framework. The formalism is explained in detail while the methodology is described in the rest of this chapter 3.

3.3 Generating Big Data

This section explains the step that generates big data using LVC simulations. First, a conceptual model is developed from the real world by the expert system which should be composed of more than three experts according to the analysis scale. The conceptual model includes the potential MOEs and MOPs, input and output data, and scenarios.

The potential MOEs and MOPs are recommended by the expert system. Each potential MOE is associated with more than a potential MOP, and thus, a variety of pairs exist with a potential MOE and potential MOP(s). Also, some potential MOPs can be simultaneously related to different potential MOEs as shown in Table 5. The potential MOEs and MOPs are represented by Equation (3.4) and Equation (3.5). The potential MOEs and MOPs can be added, deleted, or modified during the WCE analytics when necessary by the expert system.

Table 5. Example for Relationship between MOEs and MOPs

MOE_1	MOE_2	MOE_3
MOP_1, **MOP_2**, MOP_3	**MOP_2**, MOP_4, MOP_5, MOP_6	**MOP_1**, **MOP_2**, MOP_7

$$\boldsymbol{PME} = \{"y_1", "y_2", \dots, "y_k"\} \tag{3.4}$$

$$\boldsymbol{PMP} = \left\{ x_{k,i,j} \,\middle|\, x_{k,i,j} = \left(\boxed{\begin{matrix} x_{1,1,1} & x_{1,1,2} & \cdots & x_{1,1,j} \\ x_{1,2,1} & x_{1,2,2} & & x_{1,2,j} \\ \vdots & & \ddots & \vdots \\ x_{1,i,1} & x_{1,i,2} & \cdots & x_{1,i,j} \end{matrix}}, \dots, \boxed{\begin{matrix} x_{k,1,1} & x_{k,1,2} & \cdots & x_{k,1,j} \\ x_{k,2,1} & x_{k,2,2} & & x_{k,2,j} \\ \vdots & & \ddots & \vdots \\ x_{k,i,1} & x_{k,i,2} & \cdots & x_{k,i,j} \end{matrix}} \right) \right\}$$

(3.5)

PME: Potential MOEs Set

PMP: Potential MOPs Set

y_k: k_{th} Potential MOE

$x_{k,i,j}$: Value on i_{th} Simulation and j_{th} Potential MOP on k_{th} Potential MOE

Judgement on input and output data is an important process in the conceptual model. This is because data to analyze the potential MOE and MOPs is generated based on input and output data into a computational model of the WCE. Input data has two types which are fixed and unfixed input data. The fixed input data is not directly related to MOPs but can indirectly influence MOPs. The fixed input data is also necessary to model real-life agents. It can be changed to the potential MOPs by the expert system if necessary. The unfixed input data is directly related to the potential MOPs unlike the fixed input data. Its values within defined range are implemented in various ways to the computational model in support of, finally, identifying a relationship between each MOE and MOPs. MOPs can involve all attributes for unfixed input data or attributes for unfixed input data and combinations of unfixed input data. Unfixed input data for attributes related to the live and virtual simulation models is inputted by a weapon operator system that is composed of users who operate live and virtual simulation models. At this time, the operators are observed by a data engineer system that collects additional data related to the potential MOPs. The additional data can be the operators' physical and behavioral human factor data which are classified as unstructured data. Unfixed input data for attributes related to the constructive simulation model is manually entered into the CGF of the computational model by an M&S engineer system. An input data set is represented as shown in Equation (3.6). The input data set has two subset matrixes which are the fixed input data set and the unfixed input data set as shown in Equation (3.7), and Equation (3.8).

ξ_{MS}, ξ_{WS}: ***PME***, ***PMP*** → ***ID*** =

$$\left\{ i_{i,j} \;\middle|\; i_{i,j} = \begin{pmatrix} i_{1,1} & i_{1,2} & \cdots & i_{1,j} \\ i_{2,1} & i_{2,2} & & i_{2,j} \\ \vdots & & \ddots & \vdots \\ i_{i,1} & i_{i,2} & \cdots & i_{i,j} \end{pmatrix} \right\} = [\boldsymbol{FD} \quad \boldsymbol{UD}] \tag{3.6}$$

$$\boldsymbol{FD} = \left\{ c_{i,l} \;\middle|\; c_{i,l} = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,l} \\ c_{2,1} & c_{2,2} & & c_{2,l} \\ \vdots & & \ddots & \vdots \\ c_{i,1} & c_{i,2} & \cdots & c_{i,l} \end{pmatrix} \wedge \boldsymbol{\xi_{MS}}\left(c_{i,l}\right) = 'True' \right\} \tag{3.7}$$

$$\boldsymbol{UD} = \left\{ h_{i,m} \;\middle|\; h_{i,m} = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,m} \\ h_{2,1} & h_{2,2} & & h_{2,m} \\ \vdots & & \ddots & \vdots \\ h_{i,1} & h_{i,2} & \cdots & h_{i,m} \end{pmatrix} \wedge \boldsymbol{\xi_{WS}}\left(h_{i,m}\right) = 'True' \right\} \tag{3.8}$$

ξ_{MS}: M&S Engineer System

ξ_{WS}: Weapon Operator System

ID: Input Data Set

FD: Fixed Input Data Set

UD: Unfixed Input Data Set

$i_{i,j}$: Input Data Value on i_{th} Simulation and j_{th} Attribute

$f_{i,l}$: Fixed Input Data Value of CGF on i_{th} Simulation and l_{th} Attribute

$u_{i,m}$: Unfixed Input Data Value on i_{th} Simulation and m_{th} Attribute

$(j = l + m)$

Attributes for the output data have to be considered to be able to explain and analyze a relationship between each MOE and MOPs. The attributes for the output data have to include all potential MOEs and MOPs. The output data adds a time dimension received from the computational model. The dimension can offer data to analyze the WCE in terms of the time series. Also, the output data does not only include the structured data but also the semi-structured and unstructured data. Equation (3.9) shows the output data set for the structured data from the input data set through the computational model. This is only for the structured data because the semi-structured and unstructured data does not have a pre-defined data format.

δ_{CM}: $\boldsymbol{ID} \rightarrow \boldsymbol{OD_s}$ =

$$\left\{ O_{i,l,m,n} \;\middle|\; O_{i,l,m,n} = \begin{pmatrix} \begin{matrix} O_{1,1,1,1} & O_{1,1,1,2} & \cdots & O_{1,1,1,n} \\ O_{1,1,2,1} & O_{1,1,2,2} & & O_{1,1,2,n} \\ \vdots & & \ddots & \vdots \\ O_{1,1,m,1} & O_{1,1,m,2} & \cdots & O_{1,1,m,n} \end{matrix}, \dots, & \begin{matrix} O_{1,l,1,1} & O_{1,l,1,2} & \cdots & O_{1,l,1,n} \\ O_{1,l,2,1} & O_{1,l,2,2} & & O_{1,l,2,n} \\ \vdots & & \ddots & \vdots \\ O_{1,l,m,1} & O_{1,l,m,2} & \cdots & O_{1,l,m,n} \end{matrix} \\ \begin{matrix} O_{2,1,1,1} & O_{2,1,1,2} & \cdots & O_{2,1,1,n} \\ O_{2,1,2,1} & O_{2,1,2,2} & & O_{2,1,2,n} \\ \vdots & & \ddots & \vdots \\ O_{2,1,m,1} & O_{2,1,m,2} & \cdots & O_{2,1,m,n} \end{matrix}, \dots, & \begin{matrix} O_{2,l,1,1} & O_{2,l,1,2} & \cdots & O_{2,l,1,n} \\ O_{2,l,2,1} & O_{2,l,2,2} & & O_{2,l,2,n} \\ \vdots & & \ddots & \vdots \\ O_{2,l,m,1} & O_{2,l,m,2} & \cdots & O_{2,l,m,n} \end{matrix} \\ & \ddots \\ \begin{matrix} O_{i,1,1,1} & O_{i,1,1,2} & \cdots & O_{i,1,1,n} \\ O_{i,1,2,1} & O_{i,1,2,2} & & O_{i,1,2,n} \\ \vdots & & \ddots & \vdots \\ O_{i,1,m,1} & O_{i,1,m,2} & \cdots & O_{i,1,m,n} \end{matrix}, \dots, & \begin{matrix} O_{i,l,1,1} & O_{i,l,1,2} & \cdots & O_{i,l,1,n} \\ O_{i,l,2,1} & O_{i,l,2,2} & & O_{i,l,2,n} \\ \vdots & & \ddots & \vdots \\ O_{i,l,m,1} & O_{i,l,m,2} & \cdots & O_{i,l,m,n} \end{matrix} \end{pmatrix} \right\} \tag{3.9}$$

δ_{CM}: Computational Model of the WCE

$\boldsymbol{OD_s}$: Output Data Set for the Structured Data

$O_{i,l,m,n}$: Value on i_{th} Simulation, l_{th} Time, m_{th} Agent, and n_{th} Attribute

A variety of scenarios are organized for simulating the weapon that is analyzed. At this time, the scenarios should be checked by the expert system to decide whether the scenarios are appropriate compared to the field manual. The scenarios are described as proper forms which are able to help avoid misunderstandings.

The computational model is developed with LVC simulation models, which are based on the HLA or/and DIS with a target federation as shown in Figure 18. The HLA based federation is used for the constructive simulations and the DIS based federation is applied to the virtual simulations.

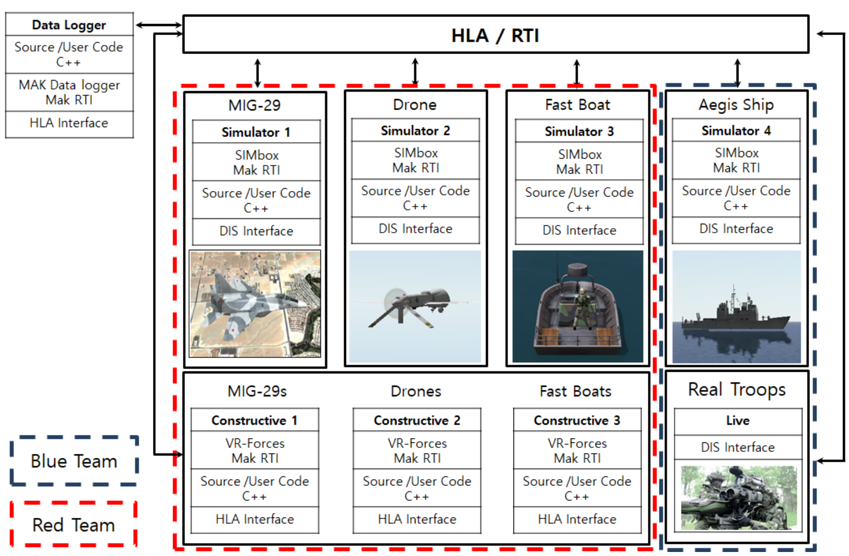

Figure 18. Example on Federation for LVC Simulations

Gamers control entities, which exist in the constructive simulation model, without their own intentions, but just by the orders of the command post. Weapon operators run the simulators which are virtual simulation models. Real troops, wearing Multiple Integrated Laser Engagement System (MILES) equipment, join the engagement. The command post, the gamers, the weapon operators, and the real troops send and receive their information from each other through communications as shown in Figure 19. Structured, semi-structured, or unstructured data, such as the log data, the voice records, the texts, the pictures, the videos, etc., can be generated through the communications in the developed computational model.

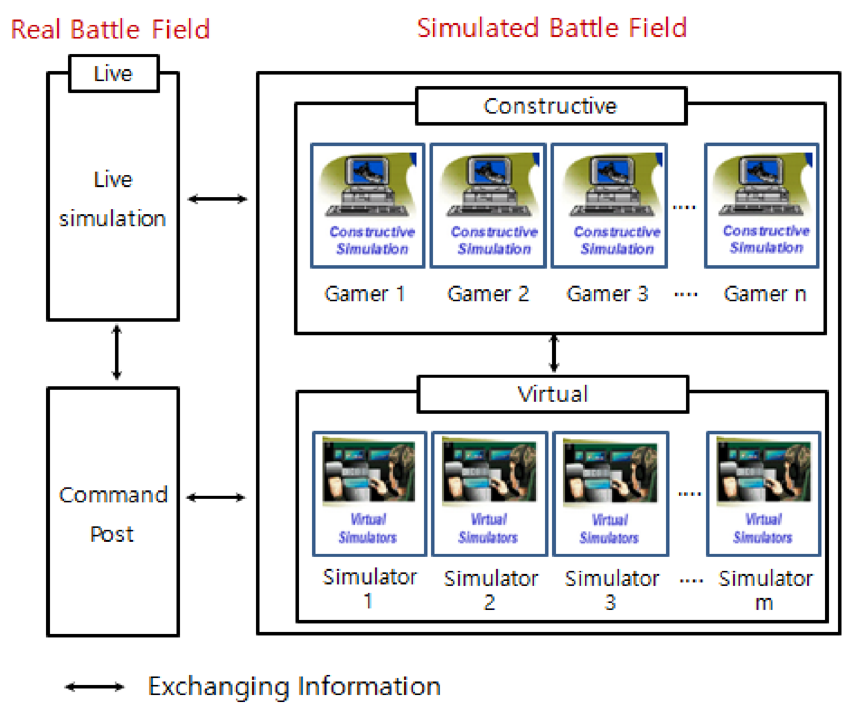

Figure 19. Information Exchange among LVC Simulation models

For the computational models a variety of combinations of simulation models can be utilized; such as Virtual and Constructive (VC) simulation models, Live and Constructive (LC) simulation models, or Live and Virtual (LV) simulation models. Eventually the LVC simulations are used for the WCE analytics in order to obtain higher fidelity. This higher fidelity is achieved by reflecting the real environment and the Human System Integration (Hsinchun et al., 2012) and using 6DOF simulation models established by the defense contractors and/or weapon manufacturers.

Input data decided in the conceptual model is implemented to the developed computational model which is LVC simulation models. The computational model is simulated by various input data which involves information on a variety of scenarios. The output data is generated by simulating the computational model.

The model to generate big data is validated and verified during the development phase as shown in the Table 6. When the conceptual model is developed, it is validated showing that the model is an accurate representation of the real world. The validation method used in WCE analytics is the reviews by Subject-Matter Experts (SMEs). When the conceptual model is implemented to the computational model, the verification is done using the animation as a dynamic testing method, which is used to trace the process in verifying if the model is built correctly or not. The computational model is validated by comparing with the real world. If the data of real world is enough to compare with data generated by the computational model, the hypothesis testing as a formal testing method is done by using the t-test. Also, the validation is done during the simulations by using the animation to check whether the model represents the real world according to the intended use. The validation is done by the SMEs' reviews which are completed in the following processes: a) identify a problem domain, b) build a review scope, method, and criteria for assessment, c) evaluate and select SMEs to review the model, d) plan the review schedule, e) review the model by SMEs, f) offer review results, and g) decide on whether the model is accepted or rejected.

Table 6. V&V for the Model

Area	Real World → Conceptual Model	Conceptual Model → Computational Model	Computational Model ↔ Real World
V&V	Validation	Verification	Validation
Method	• Informal Testing: Reviews by Subject-Matter Experts (SMEs)	• Dynamic Testing: Animation	• Formal Testing: Hypothesis Testing • Dynamic Testing: Animation • Informal Testing: Reviews by Subject-Matter Experts (SMEs)

3.4 Collecting Additional Data Related to WCE

The additional data related to WCE are collected from various sources to comprehensively analyze WCE. The additional data can be as the following: a) human input data which cannot be generated by LVC simulations, b) data stored by similar existing models, and c) other necessary data for WCE analytics.

One is the operators' physical and behavioral human factor data collected by the data engineer system. It is the data related to the potential MOPs which is collected by observing and surveying weapon operators using usability assessment methods (W. Jung, Lowe, Rabelo, Lee, & Kwon, 2017). The components of human factor data can be the stress level, observation error level, reaction level, proficiency level, gender and age, familiarity with simulators, blood alcohol level, etc. Another one is the constraints data. It is invested and extracted from various sources of the interior and exterior military; such as the number of operable weapon systems, limitation of performance of weapon systems, loadable number of system components, price information of weapon systems or components, etc. The data plays a role as constraints to solve the optimization problem. The above additional MOPs data and constraints data are transformed to the structured data as shown Equation (3.10) and Equation (3.11). The data collected from existing models and existing data in other sources can be used to support WCE analytics with the output data generated by the computational model. The collected data can be the structured, semi-structured, or unstructured data. If the data is the structured format, it is combined with the additional data set for MOPs. If not so, the semi-structured and unstructured format is kept. All additional data is needed for the data engineer system and the stakeholder system to check whether the collected data is proper to analyze WCE related to proceeding research. Figure 20 shows a type of the additional data as mentioned above.

$\boldsymbol{\theta}_{SS}$, $\xi_{DS} \rightarrow \boldsymbol{AD_a}$ =

$$\left\{ ad_{a_{i,j}} \middle| ad_{a_{i,j}} = \begin{pmatrix} ad_{a_{1,1}} & ad_{a_{1,2}} & \cdots & ad_{a_{1,j}} \\ ad_{a_{2,1}} & ad_{a_{2,2}} & & ad_{a_{2,j}} \\ \vdots & & \ddots & \vdots \\ ad_{a_{i,1}} & ad_{a_{i,2}} & \cdots & ad_{a_{i,j}} \end{pmatrix} \wedge \boldsymbol{\theta_{SS}}\,(ad_a), \boldsymbol{\xi_{DS}}\,(ad_a) = 'True' \right\} \tag{3.10}$$

$\boldsymbol{\theta}_{SS}$, $\xi_{DS} \rightarrow \boldsymbol{AD_b}$=

$$\left\{ \left(ad_{b_i}, \ldots, ad_{b_j}\right) \middle| \left(ad_{b_i}, \ldots, ad_{b_j}\right) \in \boldsymbol{PMP} \wedge \boldsymbol{\theta_{SS}}\,(ad_b), \boldsymbol{\xi_{DS}}\,(ad_b) = 'True' \right\} \tag{3.11}$$

$\boldsymbol{\theta}_{SS}$: Stakeholder System

$\boldsymbol{\xi_{DS}}$: Data Engineer System

$\boldsymbol{AD_a}$: Additional Data Set for MOPs

$\boldsymbol{AD_b}$: Additional Data Set for Constraints

PMP: Potential MOPs Set

$ad_{a_{i,j}}$: Additional Data Value for MOPs on i_{th} Simulation and j_{th} Attribute

ad_{b_i}: Additional Data Value for Constraints on i_{th} Attribute

ad_{b_j}: Additional Data Value for Constraints on j_{th} Attribute

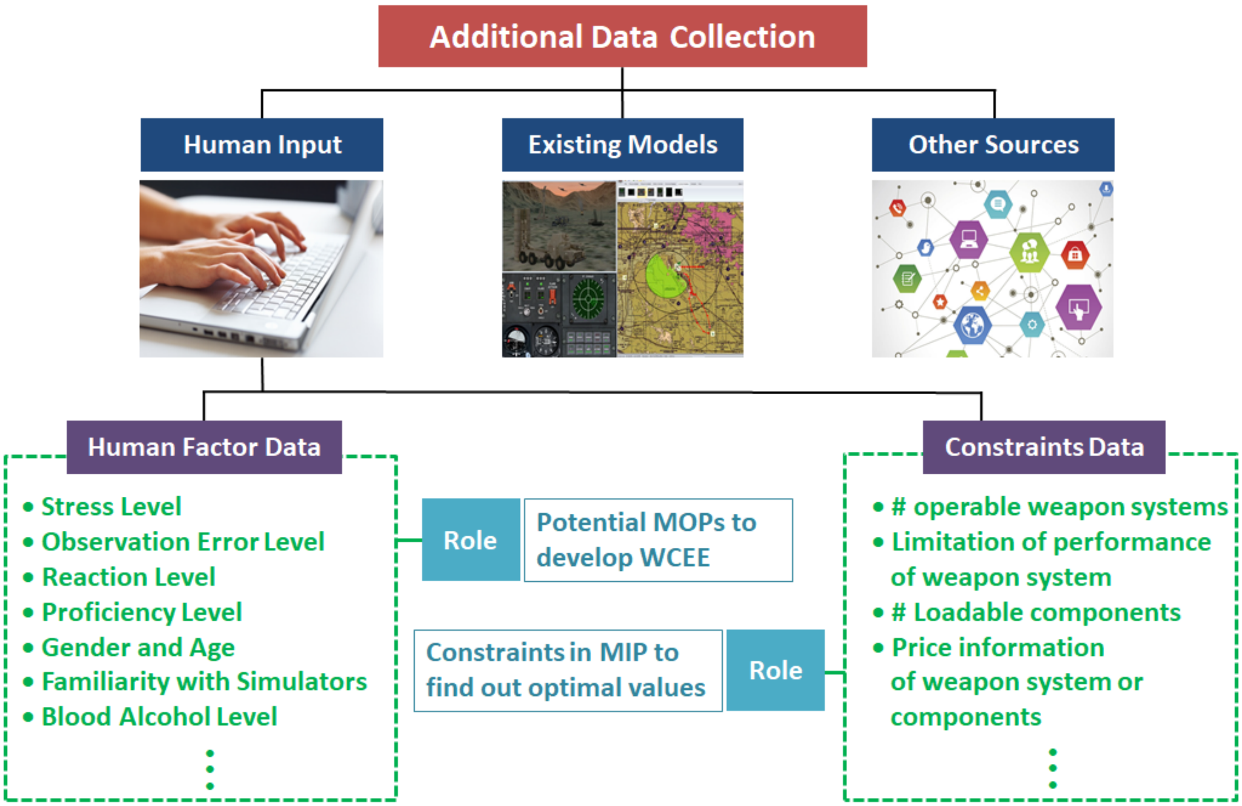

Figure 20. Example on Additional Data Collection

The data collected by other sources is validated by identifying whether the data is valid, reasonable, and sensible before processing big data. SMEs do this process by checking who collected the data, when the data was collected, how the data was collected, where the data was collected, what data was collected, etc.

3.5 Processing Big Data for WCE Analytics

This section consists of three processes as the following: a) selection, b) cleaning, integration, and transformation, and c) storage and extraction.

First, the data, which is related to the WCE analytics, is selected from the output data set for the structured data and the additional data set by the expert system as shown in Equation (3.12). The expert system does not only select the primary data based on the potential MOE and MOPs, but also the secondary data to identify the MOPs influencing WCE for the future works.

$\boldsymbol{\theta_{ES}}$: $\boldsymbol{OD_s}, \boldsymbol{AD_a}, \boldsymbol{AD_b} \rightarrow \boldsymbol{SD} =$

$$\left\{sd_{\,p,q,r,s} \,\middle|\, sd_{p,q,r,s} \in (\boldsymbol{OD_s} \cup \boldsymbol{AD_a} \cup \boldsymbol{AD_b}) \land \boldsymbol{\theta_{ES}}\,(sd) = \text{'}True\text{'}\right\} \qquad (3.12)$$

$$sd_{p,q,r,s} = \begin{pmatrix} \begin{matrix} sd_{1,1,1,1} & sd_{1,1,1,2} & \cdots & sd_{1,1,1,s} \\ sd_{1,1,2,1} & sd_{1,1,2,2} & & sd_{1,1,2,s} \\ \vdots & & \ddots & \vdots \\ sd_{1,1,r,1} & sd_{1,1,r,2} & \cdots & sd_{1,1,r,s} \end{matrix} & ,\cdots, & \begin{matrix} sd_{1,q,1,1} & sd_{1,q,1,2} & \cdots & sd_{1,q,1,s} \\ sd_{1,q,2,1} & sd_{1,q,2,2} & & sd_{1,q,2,s} \\ \vdots & & \ddots & \vdots \\ sd_{1,q,r,1} & sd_{1,q,r,2} & \cdots & sd_{1,q,r,s} \end{matrix} \\ \begin{matrix} sd_{2,1,1,1} & sd_{2,1,1,2} & \cdots & sd_{2,1,1,s} \\ sd_{2,1,2,1} & sd_{2,1,2,2} & & sd_{2,1,2,s} \\ \vdots & & \ddots & \vdots \\ sd_{2,1,r,1} & sd_{2,1,r,2} & \cdots & sd_{2,1,r,s} \end{matrix} & ,\cdots, & \begin{matrix} sd_{2,q,1,1} & sd_{2,q,1,2} & \cdots & sd_{2,q,1,s} \\ sd_{2,q,2,1} & sd_{2,q,2,2} & & sd_{2,q,2,s} \\ \vdots & & \ddots & \vdots \\ sd_{2,q,r,1} & sd_{2,q,r,2} & \cdots & sd_{2,q,r,s} \end{matrix} \\ & \ddots & \\ \begin{matrix} sd_{p,1,1,1} & sd_{p,1,1,2} & \cdots & sd_{p,1,1,s} \\ sd_{p,1,2,1} & sd_{p,1,2,2} & & sd_{p,1,2,s} \\ \vdots & & \ddots & \vdots \\ sd_{p,1,r,1} & sd_{p,1,r,2} & \cdots & sd_{p,1,r,s} \end{matrix} & ,\cdots, & \begin{matrix} sd_{p,q,1,1} & sd_{p,q,1,2} & \cdots & sd_{p,q,1,s} \\ sd_{p,q,2,1} & sd_{p,q,2,2} & & sd_{p,q,2,s} \\ \vdots & & \ddots & \vdots \\ sd_{p,q,r,1} & sd_{p,q,r,2} & \cdots & sd_{p,q,r,s} \end{matrix} \end{pmatrix}$$

$\boldsymbol{\theta_{ES}}$: Expert System

$\boldsymbol{OD_s}$: Output Data Set for the Structured Data

$\boldsymbol{AD_a}$: Additional Data Set for MOPs

$\boldsymbol{AD_b}$: Additional Data Set for Constraints

$\boldsymbol{SD}$: Selected Data Set

$sd_{p,q,r,s}$: Value on p_{th} Simulation, q_{th} Time, r_{th} Agent, and s_{th} Attribute

Second, the data is cleaned, integrated, and transformed to the unified structured data. This is named the data preparation process. Noise and inconsistent information exists in the data generated by the computational model, and in the additional data collected by the data engineer system and the stakeholder system. Therefore, it has to be cleaned for the data's reliability by removing noise and inconsistency. The structured data separated in various data sources is combined in support of effective and efficient analytics. The combined data is transformed into the proper form for reliable analytics by standardization.

Third, the data extracted by the data preparation process is named the total data set which consists of the unified structured, semi-structured, and unstructured data. The structured data of the total data set is as shown in Equation (3.13). The total data set is distributed and loaded into the storage provided by Hadoop Distributed File Systems (HDFS) that deals with storing data files across the cluster system. This is an essential step to process these big data sets for more precise WCE analytics. Also, the stored data is extracted by the MapReduce process which is responsible for the parallel processing of big data. The data extracted by the process is named the reduced data as shown in Equation (3.14). The MapReduce process has the function to reduce big data according to the given conditions. The conditions are used to extract the data which is needed to estimate WCEE. Figure 21 shows processing big data to reduced data for WCE analytics.

$\boldsymbol{\psi_{DP}}$: $\boldsymbol{SD \rightarrow TD_s}$ =

$$\left\{td_{p,q,r,s} \mid td_{p,q,r,s} \wedge \boldsymbol{\psi_{DP}}\,(td) \;= \text{'}True\text{'}\right\} \tag{3.13}$$

$$td_{p,q,r,s} = \begin{pmatrix} \begin{matrix} td_{1,1,1,1} & td_{1,1,1,2} & \cdots & td_{1,1,1,s} \\ td_{1,1,2,1} & td_{1,1,2,2} & & td_{1,1,2,s} \\ \vdots & & \ddots & \vdots \\ td_{1,1,r,1} & td_{1,1,r,2} & \cdots & td_{1,1,r,s} \end{matrix} & ,\dots, & \begin{matrix} td_{1,q,1,1} & td_{1,q,1,2} & \cdots & td_{1,q,1,s} \\ td_{1,q,2,1} & td_{1,q,2,2} & & td_{1,q,2,s} \\ \vdots & & \ddots & \vdots \\ td_{1,q,r,1} & td_{1,q,r,2} & \cdots & td_{1,q,r,s} \end{matrix} \\ \begin{matrix} td_{2,1,1,1} & td_{2,1,1,2} & \cdots & td_{2,1,1,s} \\ td_{2,1,2,1} & td_{2,1,2,2} & & td_{2,1,2,s} \\ \vdots & & \ddots & \vdots \\ td_{2,1,r,1} & td_{2,1,r,2} & \cdots & td_{2,1,r,s} \end{matrix} & ,\dots, & \begin{matrix} td_{2,q,1,1} & td_{2,q,1,2} & \cdots & td_{2,q,1,s} \\ td_{2,q,2,1} & td_{2,q,2,2} & & td_{2,q,2,s} \\ \vdots & & \ddots & \vdots \\ td_{2,q,r,1} & td_{2,q,r,2} & \cdots & td_{2,q,r,s} \end{matrix} \\ & \ddots & \\ \begin{matrix} td_{p,1,1,1} & td_{p,1,1,2} & \cdots & td_{p,1,1,s} \\ td_{p,1,2,1} & td_{p,1,2,2} & & td_{p,1,2,s} \\ \vdots & & \ddots & \vdots \\ td_{p,1,r,1} & td_{p,1,r,2} & \cdots & td_{p,1,r,s} \end{matrix} & ,\dots, & \begin{matrix} td_{p,q,1,1} & td_{p,q,1,2} & \cdots & td_{p,q,1,s} \\ td_{p,q,2,1} & td_{p,q,2,2} & & td_{p,q,2,s} \\ \vdots & & \ddots & \vdots \\ td_{p,q,r,1} & td_{p,q,r,2} & \cdots & td_{p,q,r,s} \end{matrix} \end{pmatrix}$$

$\boldsymbol{\psi_{DP}}$: Data Preparation Process

$\boldsymbol{TD_s}$: Total Data Set for the Structured Data

$td_{p,q,r,s}$: Value on p_{th} Simulation, q_{th} Time, r_{th} Agent, and s_{th} Attribute

ψ_{MP}: $\boldsymbol{TD_s} \rightarrow \boldsymbol{RD}$ =

$$\{rd_{k,p,q} \mid rd_{k,p,q} \in \boldsymbol{TD} \wedge \boldsymbol{\psi_{MR}}\,(rd) = \text{‘}True\text{’}\} \tag{3.14}$$

$$rd_{k,p,q} = \begin{pmatrix} \begin{matrix} rd_{1,1,1} & rd_{1,1,2} & \cdots & rd_{1,1,q} \\ rd_{1,2,1} & rd_{1,2,2} & & rd_{1,2,q} \\ \vdots & & \ddots & \vdots \\ rd_{1,p,1} & rd_{1,p,2} & \cdots & rd_{1,p,q} \end{matrix} & ,\dots, & \begin{matrix} rd_{k,1,1} & rd_{k,1,2} & \cdots & rd_{k,1,q} \\ rd_{k,2,1} & rd_{k,2,2} & & rd_{k,2,q} \\ \vdots & & \ddots & \vdots \\ rd_{k,p,1} & rd_{k,p,2} & \cdots & rd_{k,p,q} \end{matrix} \end{pmatrix}$$

$\boldsymbol{\psi_{MR}}$: MapReduce Process

RD: Reduced Data Set

$\boldsymbol{TD_s}$: Total Data Set for the Structured Data

$rd_{k,p,q}$: Value on p_{th} Simulation and q_{th} Attribute (q=1: MOE, q=2,…,q: MOP) in a Subset including k_{th} MOE

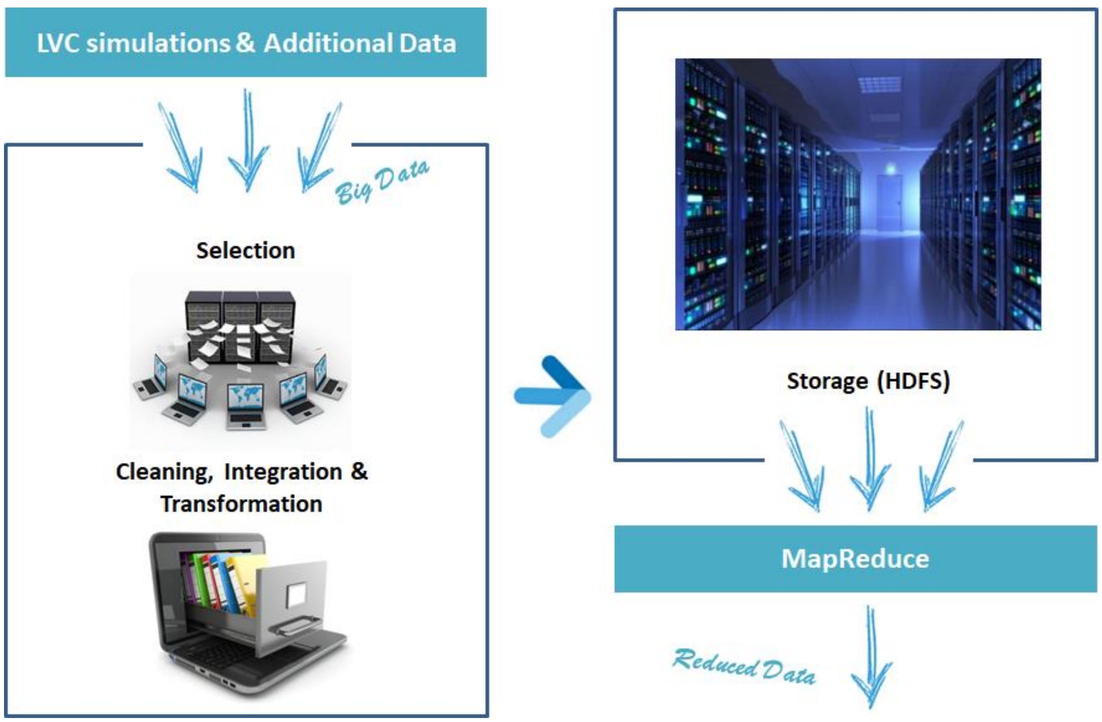

Figure 21. Processing Big Data to Reduced Data for WCE Analytics

The selected data set and total data set are validated for keeping the accuracy and the consistency of data which flows from a variety of sources during processing big data by SMEs. The reduced data set is validated by SMEs and verified by the debugging the code during the preparation and MapReduce processes. The V&V for processing big data is shown in Table 7.

Table 7. V&V for Processing Big Data

Area	Output Data & Additional Data → Selected Data → Total Data	Total Data → Reduced Data	
V&V	Validation	verification	Validation
Method	• Informal Testing: Reviews by Subject-Matter Experts (SMEs)	• Static Testing: Debugging the Code	• Informal Testing: Reviews by Subject-Matter Experts (SMEs)

3.6 Estimating WCE Equation & Optimal Values

This section focuses on estimating the WCE equation and optimal values of factors influencing WCE. The computational model is stochastic. Although the computational model is simulated under the same scenario, the simulation results are various. Therefore, the number of simulations per each scenario is different in the reduced data set. This means that each scenario has a different leverage unless the reduced data set is normalized to the same number of simulations per each scenario for the potential. The normalization process is applied to solve this problem. It is that the expected value of each MOE is calculated per a scenario and MOPs on each MOE are used per a scenario. The reduced data is normalized as shown in Equation (3.15), Equation (3.16), and Equation (3.17).

$\boldsymbol{\psi_{NP}}$: $\boldsymbol{RD} \rightarrow \boldsymbol{ND} = [\boldsymbol{ND_a}\ \ \boldsymbol{ND_b}] =$

$$\left\{ nd_{k,i,j} \middle| nd_{k,i,j} = \begin{pmatrix} nd_{1,1,1} & nd_{1,1,2} & \cdots & nd_{1,1,m} \\ nd_{1,2,1} & nd_{1,2,2} & & nd_{1,2,m} \\ \vdots & & \ddots & \vdots \\ nd_{1,i,1} & nd_{1,i,2} & \cdots & nd_{1,i,m} \end{pmatrix}, \ldots, \begin{pmatrix} nd_{k,1,1} & nd_{k,1,2} & \cdots & nd_{k,1,n} \\ nd_{k,1,1} & nd_{k,2,2} & & nd_{k,2,n} \\ \vdots & & \ddots & \vdots \\ nd_{k,i,1} & nd_{k,i,2} & \cdots & nd_{k,i,n} \end{pmatrix} \right\}$$

(3.15)

$$\boldsymbol{ND_a} = \left\{ nd_{k,i,1} \middle| \, nd_{k,i,1} = (rd_{k,p,1} | rd_{k,p,1} \in \boldsymbol{RD} \wedge \boldsymbol{\psi_{NP}}(rd_{k,p,1}) =' True') \right\} \quad (3.16)$$

$$\boldsymbol{ND_b} = \left\{ nd_{k,i,j\neq 1} \middle| \, nd_{k,i,j\neq 1} = (rd_{k,p,q\neq 1} \middle| rd_{k,p,q\neq 1} \in \boldsymbol{RD} \wedge \boldsymbol{\psi_{NP}}(rd_{k,p,q\neq 1}) = 'True') \right\}$$

(3.17)

$\boldsymbol{\psi_{NP}}$: Normalization Process

ND: Normalized Data Set

$\boldsymbol{ND_a}$: Normalized Data Set for MOEs

$\boldsymbol{ND_b}$: Normalized Data Set for MOPs

RD: Reduced Data Set

$nd_{k,i,j}$: Value on i_{th} Scenario and j_{th} Attribute (j=1: MOE, j=2,…,m or n: MOP) in a Subset including k_{th} MOE

$rd_{k,p,q}$: Value on p_{th} Simulation and q_{th} Attribute (q=1: MOE, q=2,…,q: MOP) in a Subset including k_{th} MOE

An algorithm flowchart is developed for estimating WCEE based on various regression methods. The algorithm flowchart is divided by four phases which are the MOPs selection, data attribute analysis, WCEEs development, and WCEE selection as shown in the Figure 22. The process is the following: a) MOPs selection is the MOPs influencing each MOE are statistically extracted from the ***ND***, b) Data attribute analysis is executed to distinguish whether multi-collinearity, outlier and heteroscedasticity exist or not, c) WCEEs development is to build proper WCEEs generated by regression methods using the statistical evaluation, and d) WCEE selection is a process to decide the best alternative based on the model accuracy. The MOPs' selection phase is the process to determine the influence of the MOPs on each MOE by comparing Akaike information criterion (AIC) with a stepwise selection method. When a MOPs' combination has the lowest AIC, it is decided for estimating WCEE.

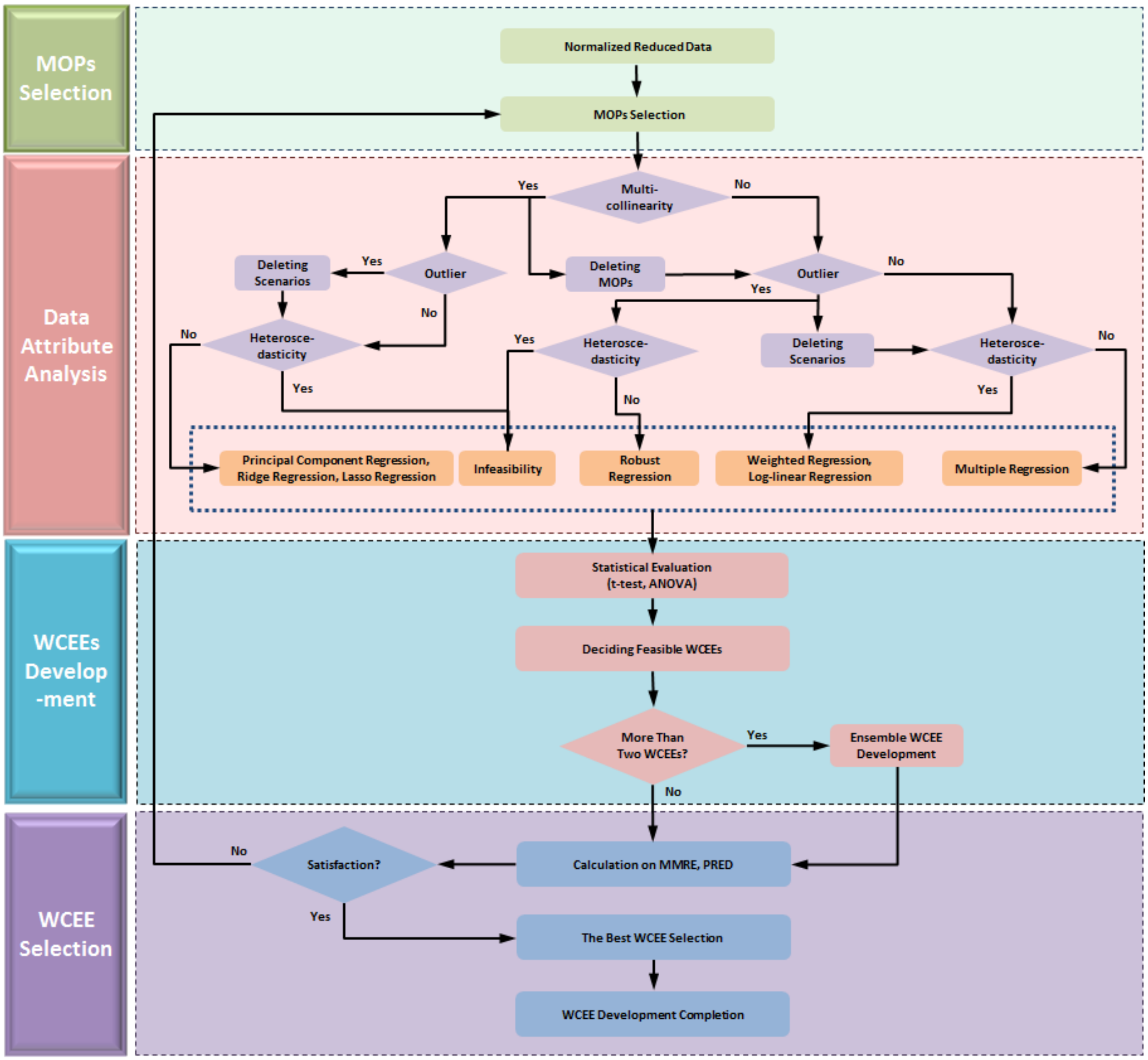

Figure 22. WCEE Development Algorithm Flowchart using Various Regression Methods

The data attribute analysis phase is executed through three layers such as the following: a) a multi-collinearity check layer, b) an outlier check layer, and c) a heteroscedasticity check layer. The multi-collinearity is checked using Variation Inflation Factor (VIF) to see if a MOP can be linearly related to the others with a considerable degree in the first layer. The outlier is checked using the studentized residual and the studentized deleted residual to see if a value generated by a simulation is distant from other values in the second layer. The heteroscedasticity is checked using the residual analysis to see if a collection of MOPs has subsets that have different variabilities from other subsets in the third layer. The identified problems can be solved by using an appropriate regression method based on the data attributes.

In WCEEs development process phase, the WCEEs are established by various regression methods and then are evaluated to decide whether the WCEEs are statistically meaningful models through the t-test and ANalysis Of Variance (Weatherly et al., 1996). The estimated coefficients of MOPs are statistically evaluated by the t-test as well as the entire WCEE is by the ANOVA at an alpha level of 0.05. If there are more than two WCEEs which satisfy the criteria, each WCEE has weight according to the accuracy index. The weights are calculated based on the accuracy index as shown in Equation (3.18). The WCEEs with their own weights are linearly combined to one WCEE which can be the other prediction model other than the criteria-satisfied WCEEs. The linearly combined WCEE is named the ensemble WCEE.

$$\mathrm{W}_k = \frac{\mathrm{AI}_k}{\sum_{k=1}^{l} \mathrm{AI}_k} \tag{3.18}$$

W_k: k_{th} WCEE Weight

AI_k: k_{th} WCEE Accuracy Index

l: Number of WCEEs Satisfy the Criteria

In the WCEE selection phase, the best WCEE is selected through Mean Magnitude of Relative Error (MMRE) and Prediction (PRED) as shown in Equation (3.19) and Equation (3.20). The MMRE and PRED are widely used as model accuracy estimators. Conte, Dunsmore, and Shen (1986) consider MMRE $\leq$ 0.25 as an acceptable level for effort prediction models. The PRED(.25) is applied, but however, some research also uses PRED(.30). Typically PRED(.30) $\geq$ .75 is considered an acceptable model accuracy (Boehm et al., 2000). Therefore, when a WCEE satisfies criteria which are MMRE $\leq$ 0.25 and PRED(.30) $\geq$ .75, the prediction model is statistically reliable. If there are more than two WCEEs which satisfy the criteria, the best WCEE is selected of the acceptable WCEEs including the ensemble model based on their accuracy. If only a WCEE satisfies the criteria, the WCEE is finally selected. If there is not a WCEE satisfying the criteria, MOPs are selected again.

$$\text{MMRE} = \frac{1}{n} \times \text{MRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{3.19}$$

$$\text{PRED}(x) = \frac{1}{n}\sum_{i=1}^{n}\begin{cases}1 & if\ MRE_i \leq 1\\ 0 & otherwise\end{cases} \tag{3.20}$$

n: Number of Data Points

y_i: y Value of i_{th} Data Point

$\hat{y}_i$: Estimated y Value of i_{th} Data Point

k: Number of Data Points Satisfying MRE $\leq x$

The final WCEEs are composed of the estimated MOEs' values set, the MOPs' values set, and the estimated coefficients on MOPs as shown in Equation (3.21), Equation (3.22), Equation (3.23), and Equation (3.24).

$\boldsymbol{\lambda}_{WA}$: ***ND*** $\rightarrow$ ***WCEE***:

$$\widehat{\boldsymbol{Y}} = \hat{\beta}\boldsymbol{X} \tag{3.21}$$

$$\widehat{\boldsymbol{Y}} = \left\{ \hat{y}_{k,i} \;\middle|\; \hat{y}_{k,i} = \left. \begin{matrix} \hat{y}_{1,1} & \hat{y}_{1,2} & \cdots & \hat{y}_{1,k} \\ \hat{y}_{2,1} & \hat{y}_{2,2} & & \hat{y}_{2,k} \\ \vdots & & \ddots & \vdots \\ \hat{y}_{i,1} & \hat{y}_{i,2} & \cdots & \hat{y}_{i,k} \end{matrix} \right) \right\} \tag{3.22}$$

$$\boldsymbol{X} = \left\{ x_{k,i,j} \;\middle|\; x_{k,i,j} = \left(\boxed{\begin{matrix} 1 & x_{1,1,1} & x_{1,1,2} & \cdots & x_{1,1,j} \\ 1 & x_{1,2,1} & x_{1,2,2} & & x_{1,2,j} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{1,i,1} & x_{1,i,2} & \cdots & x_{1,i,j} \end{matrix}}, \ldots, \boxed{\begin{matrix} 1 & x_{k,1,1} & x_{k,1,2} & \cdots & x_{k,1,j} \\ 1 & x_{k,2,1} & x_{k,2,2} & & x_{k,2,j} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{k,i,1} & x_{k,i,2} & \cdots & x_{k,i,j} \end{matrix}} \right) \right\} \tag{3.23}$$

$$\widehat{\boldsymbol{\beta}} = \left\{ \hat{\beta}_{j,k} \;\middle|\; \hat{\beta}_{j,k} = \left. \begin{matrix} \hat{\beta}_{0,1} & \hat{\beta}_{0,2} & \cdots & \hat{\beta}_{0,k} \\ \hat{\beta}_{1,1} & \hat{\beta}_{1,2} & & \hat{\beta}_{1,k} \\ \vdots & & \ddots & \vdots \\ \hat{\beta}_{j,1} & \hat{\beta}_{j,2} & \cdots & \hat{\beta}_{j,k} \end{matrix} \right) \right\} \tag{3.24}$$

$\boldsymbol{\lambda}_{WA}$: WCEE Development Algorithm

ND: Normalized Data

WCEE: Weapon Combat Effectiveness Equation

$\widehat{\boldsymbol{Y}}$: Estimated MOEs' Values Set

X: MOPs' Values Set

$\widehat{\boldsymbol{\beta}}$: Estimated Coefficients on MOPs Set

$\hat{y}_{k,i}$: Estimated Value on k_{th} MOE and i_{th} Simulation

$x_{k,i,j}$: Value on i_{th} Simulation and j_{th} MOP on k_{th} MOE

$\hat{\beta}_{j,k}$: Estimated Coefficient on j_{th} MOP on k_{th} MOE

$\hat{\beta}_{0,k}$: k_{th} MOE's Intercept

The optimal MOPs' values are estimated to maximize each MOE as the Mixed Integer Programming (MIP) problem. The best accurate WCEE is used as an objective function and the constraints are extracted from the conditional data set as shown in Figure 23. The optimal values are shown in Equation (3.25).

Maximize $\hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X}$ *Objective Function*

Subject to $\boldsymbol{a} \leq \boldsymbol{AX} \leq \boldsymbol{b}$ *Constraints*

$\boldsymbol{a} \geq 0, \boldsymbol{X} \geq 0$

$\boldsymbol{X} \in \mathbb{Z}^n$

$\hat{\boldsymbol{\beta}}$: *Effect vector of MOP*

$\boldsymbol{X}$: *MOP*

$\boldsymbol{A}$: *Constraint matrix of MOP*

$\boldsymbol{a}$: *Upper bound vector*

$\boldsymbol{b}$: *Lower bound vector*

Minimize $\hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X}$ *Objective Function*

Subject to $\boldsymbol{a} \leq \boldsymbol{AX} \leq \boldsymbol{b}$ *Constraints*

$\boldsymbol{a} \geq 0, \boldsymbol{X} \geq 0$

$\boldsymbol{X} \in \mathbb{Z}^n$

Figure 23. Optimal Values' Estimation using Mixed Integer Programming (MIP)

ψ_{MIP}: $\boldsymbol{WCEE}, \boldsymbol{AD_b} \rightarrow \boldsymbol{V}^{*}=$

$$\begin{bmatrix} y^* \\ x_1{}^* \\ x_2{}^* \\ \vdots \\ x_j{}^* \end{bmatrix} = \begin{bmatrix} \mathrm{a}_0 \\ \mathrm{a}_1 \\ \mathrm{a}_2 \\ \vdots \\ \mathrm{a}_j \end{bmatrix} \tag{3.25}$$

ψ_{MIP}: Mixed Integer Programing Process

WCEE: Weapon Combat Effectiveness Equation

$\boldsymbol{AD_b}$: Additional Data Set for Constraints

$\boldsymbol{V}^*$: Optimal Values' Set

The data V&V is processed in this section as shown in Table 8. The SMEs validate the data normalized from the reduced data set by the normalized process. Also, at this time it is verified that the WCEE development algorithm is correctly implemented in a statistics package program by debugging the code. It is validated by using the hypothesis testing and prediction testing that the developed WCEEs are the correct models. The validated WCEEs mean that they can be used as the prediction model for the WCE estimation. The method of debugging the code is used for the verification for the optimal value estimation. Its validation is done by using the data analysis which is the static testing method.

Table 8. V&V for Data

Area	Reduced Data → Normalized Data	Normalized Data → WCEE Development & WCEE Development → Optimal Values' Estimation	
V&V	Validation	Verification	Validation
Method	• Informal Testing: Reviews by Subject-Matter Experts	• Static Testing: Debugging the Code Traceability Assessment	• Formal Testing: Hypothesis Testing, Prediction Testing, Data Analysis

3.7 Reporting Results

The final step is to present the findings in usable knowledge formats such as with a probability map and effectiveness index to help the decision makers make decisions effectively and efficiently. Also, the document gives the decision makers comprehensive analytics results.

The document could give the decision makers beneficial information, with which they could comprehensively review analytics results and decide the best one of the various alternatives.

The probability map is composed of X, Y, and Z axes as a result of analyzing big data as shown in Table 9. At this time, X and Y axes represent MOPs of WCEE for each Blue and Red team. Also, each MOP consists of various scales. Z axis is the MOE based on X and Y. The most important MOP and its value could be visually identified by the probability map, which could help the commanders decide their alternatives. For example, when X_n of Blue team MOPs is 1.2 as shown in Table 9 while, most of MOE's values are at a high level regardless of Red team MOPs.

Table 9. Probability Map Example

MOE 0~25% 26~50% 51~75% 76~100%

			Blue Team Factors									
			X_1		X_3				…	X_n		
			0	1	0.1	0.5	0.9	1.3	…	1.2	…	4.8
Red Team Factors	X_2	0.2	48%	99%	13%	18%	23%	50%	…	100%	…	38%
		0.6	13%	18%	53%	59%	71%	99%	…	80%	…	21%
		1.0	0%	15%	99%	81%	76%	55%	…	75%	…	11%
	X_4	12	3%	67%	54%	64%	74%	87%	…	76%	…	32%
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	…	⋮
		28	34%	78%	59%	73%	79%	92%	…	94%	…	53%
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	…	⋮
	X_m	1.4	15%	19%	57%	61%	73%	98%	…	82%	…	28%
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	…	⋮
		3.6	5%	69%	54%	68%	84%	97%	…	86%	…	42%

The effectiveness index is the priority number of combinations of the MOPs' values in the order of the estimated MOE maximization in descending order as shown in Table 10. This index enables decision makers to identify the priority of the combination of the weapon system factors. It also help the decision makers get the best alternative solution under the given conditions.

Table 10. Effectiveness Index Example

MOE 0~25% 26~50% 51~75% 76~100%

Index	MOE	Blue Team Factors				Red Team Factors			
		X_1	X_3	⋯	X_n	X_2	X_4	⋯	X_m
1	100%	0	0.1	⋯	1.2	0.2	12	⋯	1.4
2	100%	1	0.1	⋯	1.2	0.2	12	⋯	1.4
3	100%	0	0.5	⋯	1.2	0.2	12	⋯	1.4
4	100%	0	0.9	⋯	1.2	0.2	12	⋯	1.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	54%	1	0.1	⋯	4.8	0.6	12	⋯	3.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
56	28%	1	0.5	⋯	4.8	1.0	28	⋯	1.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	0%	0	0.9	⋯	2.4	1.0	20	⋯	1.4

The resulting report is composed of a WCE equation, optimal values, a probability map, an effectiveness index, etc. for decision makers to comprehensively understand the results of WCE. This step can be moved back to the previous steps by the decision maker. For example, scenarios can be modified in order to generate more data that can be regenerated by the data preparation subsystem, and then, WCEEs can be developed again by the analytics subsystem.

CHAPTER FOUR: F-16 COMBAT EFFECTIVENESS AGAINST SAM

This chapter shows a case study on F-16s CE analytics against SA-8 TELARs (SAMs) using the ULM built by the suggested WCE analytics methodology. A pilot model is developed and analyzed for the case study.

4.1 Overview

This case study is conducted in two sections which are a) the data preparation for the F-16 CE analytics by LLM_{DP}, and b) F-16 CE analytics by LLM_{AN} and the visualization as shown in Figure 24. In the first section a-1) Big data is generated by VC simulations, a-2) F-16 CE-related additional data is collected, and a-3) Generated and collected big data are ready for F-16 CE analytics. In another section b-1) the F-16 CE equation is estimated and optimal values influencing F-16 CE are identified, and b-2) the F-16 analytics results are reported using visualization.

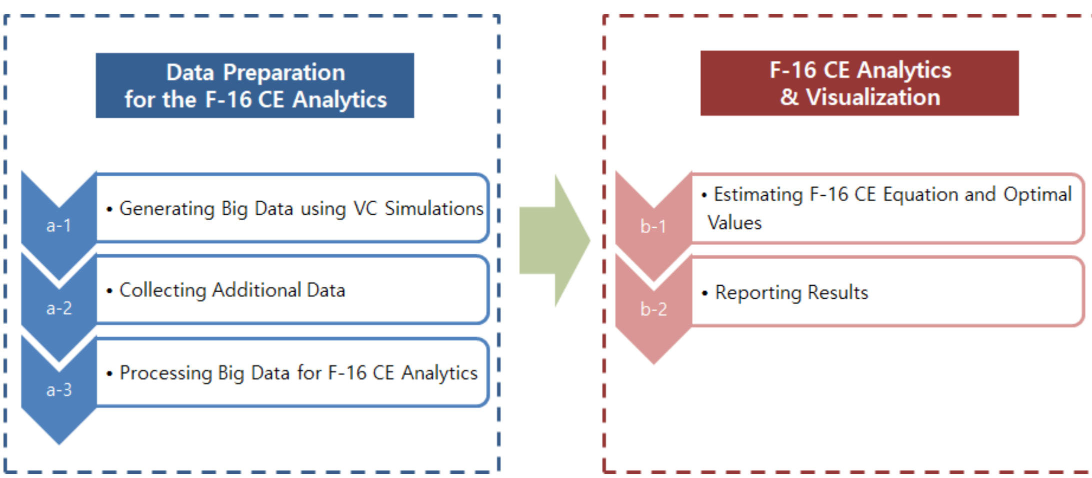

Figure 24. Case Study Process Structure for F-16 CE Analytics against SAM

4.2 Data Preparation Subsystem for F-16 CE Analytics

This section describes processes to prepare big data for F-16 CE analytics by the LLM_{DP}. These processes are completed through continuous communication between the ULM and the LLM_{DP}. The processes include the big data generation, additional data collection, and big data readiness for analytics.

4.2.1 STEP 1: Generating Big Data

This step shows how to generate big data using VC simulations which are a virtual simulation model SIMbox (http://www.simigon.com/simulation-training-development-tools) and a constructive simulation model VR-forces (https://www.mak.com/products/simulate/vr-forces). The SIMbox can be operated as a constructive simulation model according to the setting.

4.2.1.1 A conceptual model

Including the potential MOE, the potential MOPs, input and output data, and scenarios, a conceptual model is developed for the case study. The validation of the constructed concept model is done to keep the model's reliability.

4.2.1.1.1 A conceptual model construction

The conceptual model is developed by the expert system which consists of three defense experts who have served more than fifteen years either in the Air Force or Army. A potential MOE and seventeen potential MOPs are selected by the expert system for the case study. The potential MOE is the survival rate of the blue team which consists of F-16s given that the red team's core facility is destroyed. Equation (4.1) shows a set of the potential MOE.

$$\boldsymbol{PME} = \{\text{"P(FS |CD = 1)"}\} \tag{4.1}$$

PME: Potential MOEs Set

P: Probability

FS: F-16 Survival or not

CD: Core Facility Destruction $\begin{cases} 0, & Survival \\ 1, & Destruction \end{cases}$

The potential MOPs have three categories which are the weapon performance, the training profile, and the operations plan. The weapon performance includes the number of release bombs, the number of chaffs and flares, the control fire range, the control track range, the difference between control fire and track range, the control time between fire launches, and the number of missiles. They represent the physical capabilities. The training profile shows the human factors related to the WCE. The observation error and the reaction time of the human factors are applied to the training profile. The operation plan means operational alternatives to complete a mission. It includes three attack methods, the number of attack flights, three deployment methods, and the number of defense SAMs. The potential MOP data set is shown in Equation (4.2). The Table 11 explains the potential MOPs selected by the expert system.

$$\boldsymbol{PMP} = \left\{ x_{i,j} \;\middle|\; x_{i,j} = \begin{pmatrix} 6 & 30 & \cdots & 1 \\ 6 & 30 & & 1 \\ \vdots & & \ddots & \vdots \\ 5 & 70 & \cdots & 1 \end{pmatrix} \right\} \tag{4.2}$$

PMP: Potential MOPs Data Set

$x_{i,j}$: Value on i_{th} Simulation and j_{th} Potential MOP

Table 11. Potential MOPs Selected by the Expert System

MOPs		Team Types	Descriptions
Weapon Performance (7)	The number of release bombs	Blue	The number of weapons that should be released on attack
	The number of chaffs and flares	Blue	The number of loaded chaffs and flares
	Control fire range	Red	The range in miles at which to fire on the target
	Control track range	Red	The range in miles at which to start tracking the target
	Difference between control fire and track range	Red	The range difference in miles between control fire and track range
	Control time between fires	Red	Time of fire rate for only firing once
	The number of missiles	Red	The number of loaded missiles
Training Profile (2)	Observation Error	Blue	Observation error when events occur
	Reaction Time	Red	Reaction time when events occur
Operations Plan (8)	Attack method 1 (Type A)	Blue	Attack at same time as a group
	Attack method 2 (Type B)	Blue	Attack at same time as two groups which are deployed to a 90 degree direction
	Attack method 3 (Type C)	Blue	Attack at same time as two groups which are deployed to a 180 degree direction
	The number of attack flights	Blue	The number of flights to attack the enemy
	Deployment 1 (Type A)	Red	Deployment of a SAM beside the core facility
	Deployment 2 (Type B)	Red	Deployment of two SAMs in a group beside the core facility
	Deployment 3 (Type C)	Red	Deployment of two SAMs which are operated with separated positions
	The number of defense SAMs	Red	The number of SAMs to defend the enemy

There are limitations to the weapon performances, operator abilities, and operations alternatives. Since the potential MOPs' range cannot be applied to infinite values as well as simulated with unlimited values, it is necessary to define the appropriate range of the potential MOPs by the experts system. Table 12 represents values of each potential MOP which are decided by the expert system in this research. The potential MOE and MOPs are decided or considered according to results generated by VC simulations.

Table 12. Values of the potential MOPs

MOPs			Value Types	Units	Values
Weapon Performance (7)	The number of release bombs	X_1	Integer	EA	$1 \le x_1 \le 7$
	The number of chaffs and flares	X_2	Integer	EA	$10 \le x_2 \le 100$
	Control fire range	X_3	Float	Mile	$1 \le x_3 \le 10$
	Control track range	X_4	Float	Mile	$1 \le x_4 \le 20$
	Difference between control fire and track range	X_5	Float	Mile	$0 \le x_5 \le 19$
	Control time between fires	X_6	Float	Second	$5 \le x_6 \le 60$
	The number of missiles	X_7	Integer	EA	$0 \le x_7 \le 6$
Training Profile (2)	Observation Error	X_8	Float	Point	$0 \le x_8 \le 100$
	Reaction Time	X_9	Float	Point	$1 \le x_9 \le 100$
Operations Plan (8)	Attack method 1 (Type A)	X_{10}	Binary	N/A	$x_{10} = (x_{10a}, x_{10b}) = (0, 0)$
	Attack method 2 (Type B)	X_{10}	Binary	N/A	$x_{10} = (x_{10a}, x_{10b}) = (0, 1)$
	Attack method 3 (Type C)	X_{10}	Binary	N/A	$x_{10} = (x_{10a}, x_{10b}) = (1, 0)$
	The number of attack flights	X_{11}	Integer	EA	$1 \le x_{11} \le 7$
	Deployment 1 (Type A)	X_{12}	Binary	N/A	$x_{12} = (x_{12a}, x_{12b}) = (0, 0)$
	Deployment 2 (Type B)	X_{12}	Binary	N/A	$x_{12} = (x_{12a}, x_{12b}) = (0, 1)$
	Deployment 3 (Type C)	X_{12}	Binary	N/A	$x_{12} = (x_{12a}, x_{12b}) = (1, 0)$
	The number of defense SAMs	X_{13}	Integer	EA	$1 \le x_{13} \le 2$

Input and output data, which are described in the computational model, are decided for analyzing the potential MOE and MOPs. The fixed input data is selected for modeling real-life agents which are F-16, SAM, SA-8, M-117, and the core facility. The fixed input data is the default values given from the SIMbox Toolkit. Each agent has a variety of the fixed input data which belongs to the Logic Object Components (LOCs), the Console Object Components (COCs), and the Output Object Components (OOCs). Figure 25 shows an example of the fixed input data involved in three components for F-16 and SAM.

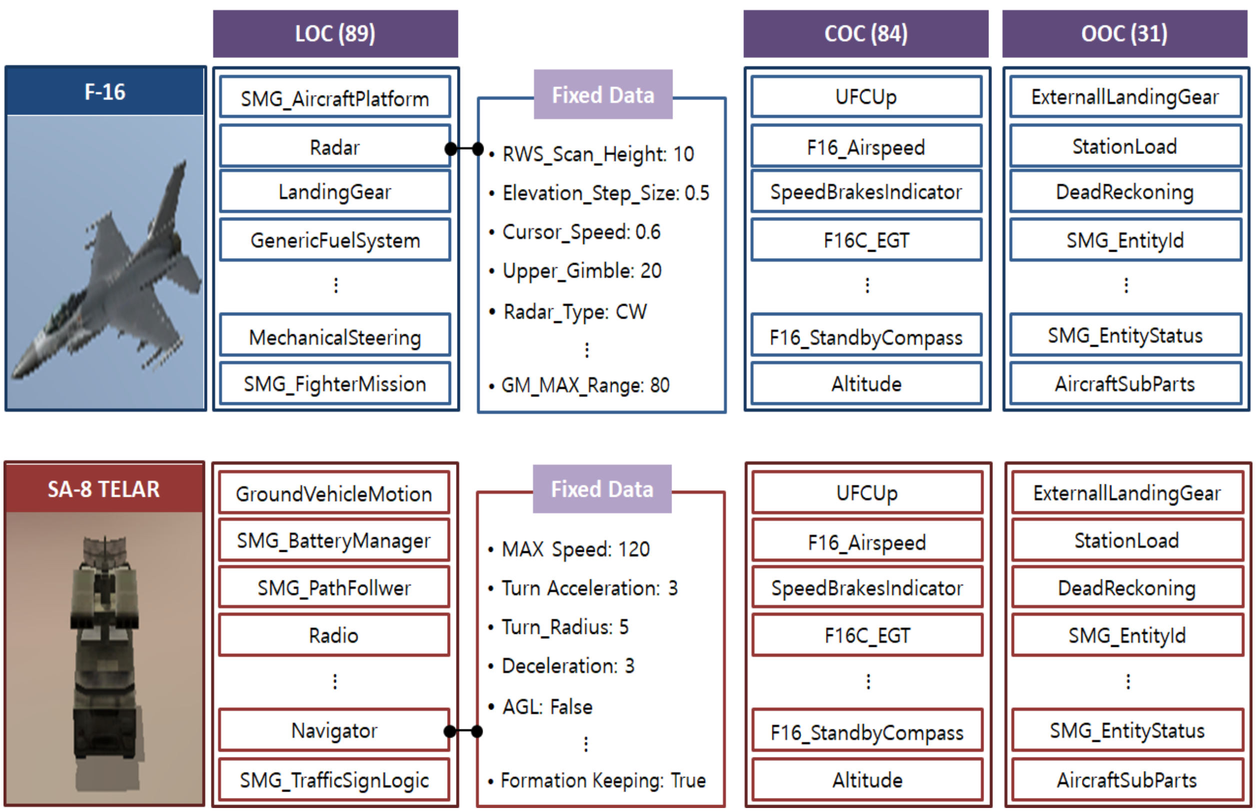

Figure 25. Main Entities’ Example on Fixed Input Data in LOC, COC, and OOC

Directly related to potential MOPs, the unfixed input data in this case study is 17 MOPs which have been discussed above. Unfixed input data for 2 attributes, which are the observation error and the reaction time happened by weapon manipulation of operators, is observed and input by the data engineer system as additional data. Unfixed input data for the other 15 attributes is manually entered into the Computer Generated Forces (CGF) of the computational model by the M&S engineer system. The input data set is composed of two subset matrixes the fixed input data set and the unfixed data set as shown in Equation (4.3), Equation (4.4), and Equation (4.5). Table 13 classifies fixed and unfixed input data attributes applied in this case study.

ξ_{MS}, ξ_{WS}: $\boldsymbol{PME}, \boldsymbol{PMP} \rightarrow \boldsymbol{ID} = [\boldsymbol{FD} \quad \boldsymbol{UD}]$

$$\left\{ i_{i,j} \;\middle|\; i_{i,j} = \begin{pmatrix} 10 & 0.5 & \cdots & 1 \\ 10 & 0.5 & & 1 \\ & \vdots & \ddots & \vdots \\ 10 & 0.5 & \cdots & 1 \end{pmatrix} \right\} \tag{4.3}$$

$$\boldsymbol{FD} = \left\{ f_{i,l} \;\middle|\; f_{i,l} = \begin{pmatrix} 10 & 0.5 & \cdots & 5 \\ 10 & 0.5 & & 5 \\ & \vdots & \ddots & \vdots \\ 10 & 0.5 & \cdots & 5 \end{pmatrix} \right\} \tag{4.4}$$

$$\boldsymbol{UD} = \left\{ u_{i,m} \;\middle|\; u_{i,m} = \begin{pmatrix} 6 & 30 & \cdots & 1 \\ 6 & 30 & & 1 \\ & \vdots & \ddots & \vdots \\ 5 & 70 & \cdots & 1 \end{pmatrix} \right\} \tag{4.5}$$

ξ_{MS}: M&S Engineer System

ξ_{WS}: Weapon Operator System

ID: Input Data Set

FD: Fixed input data Set

UD: Unfixed input data Set

$i_{i,j}$: Input Data Value on i_{th} Simulation and j_{th} Attribute

$f_{i,l}$: Fixed Input Data Value on i_{th} Simulation and l_{th} Attribute

$u_{i,m}$: Unfixed Input Data Value on i_{th} Simulation and m_{th} Attribute

$(j = l + m)$

Table 13. Attributes of Fixed and Unfixed Input Data

Categories		Attributes for Fixed input data	Attributes for Unfixed input data
CGF Input Data	Classic	▪ M-117 weight ▪ M-117 Explosion radius ▪ M-117 Explosion effect power ▪ SA-8 seeker type ▪ SA-8 wing area ▪ F-16 radar max scan range ▪ F-16 radar type ▪ SA-8 TELAR turret turn speed ⋮	▪ The number of release bombs ▪ The number of chaffs and flares ▪ Control fire range ▪ Control track range ▪ Control time between fires ▪ The number of missiles ▪ F-16 attack types (3) ▪ The number of attack flights ▪ SA-8 TELAR deployment types (3) ▪ The number of defense SAMs
	Combination	-	▪ Difference between control fire and tract range
Human Factors Input Data		-	▪ Observation Error ▪ Reaction time

The attributes for output data, which are classified as the structured data, include potential MOPs' data as well as F-16s and a core facility damage levels generated for a unit of ten milliseconds. Semi-structured and unstructured data is also generated as the output data. However, this case study does not apply semi-structured and unstructured data. The output data offers information to develop WCEE which is a relationship between a MOE and MOPs. The output data for the structured data is generated by 1,500 simulations with 67 agents during 345.000sec per a simulation.

A variety of scenarios are composed to generate big data for F-16s CE analytics against SAMs by the expert system. There are two teams which are blue and red. The blue team consists of F-16s whereas the red team is made up of SAMs and a core facility. The F-16's mission against SAMs is to destroy the red team's core facility and return to their base while surviving. Modeled to generate data for WCE analytics, agents interact with each other during engagement by trying to achieve a blue team's mission. Figure 26 shows the SysML Use Case Diagram which represents an example of the agents' interaction during an engagement between two F-16s and a SAM.

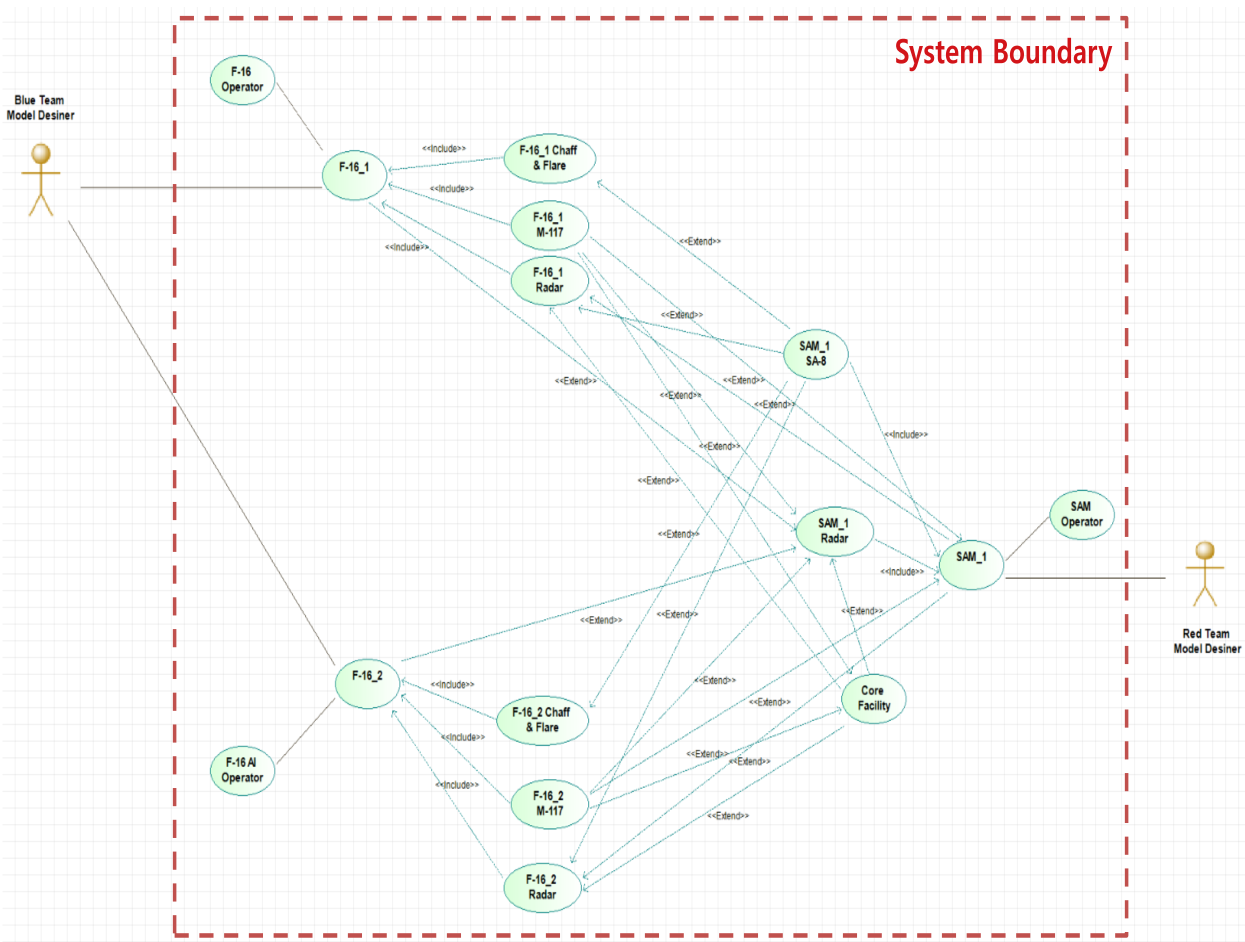

Figure 26. SysML Use Case Diagram on Agents' Interaction (e.g. Two F-16s and One SAM)

The agents' interaction occurs under various environments that are developed by reflecting various attributes and their values as shown in Figure 27. For example, one of various scenarios is that seven F-16s attack a core facility with evading SA-8s launched from two SAMs. In this scenario the F-16 has various attributes and their values which are five released bombs, one hundred chaffs and flares, the observation error automatically input by an operator, the attack method of type C. A variety of attributes and their values are applied to two SAMs for defending a core facility. They are the control fire range of ten miles, the control track range of twenty miles, the difference of the control fire and track range of ten miles, the control time between fires of ten seconds, the number of missiles of five, the reaction time automatically inserted by an operator, the deployment of type B, and the number of defense SAMs of two.

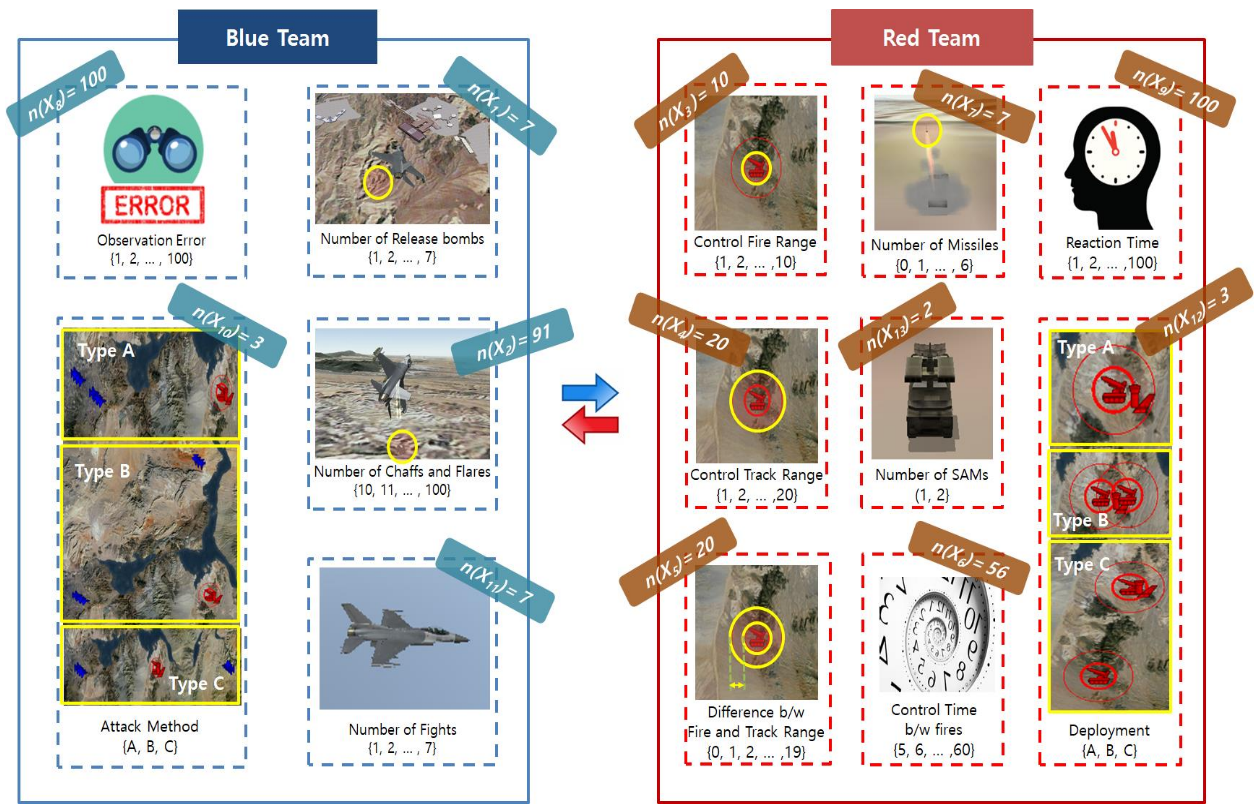

Figure 27. Construction of Scenarios

Values of a real number's type are input into discrete values which are an integer unit in support of modeling the conceptual model with finite values whereas estimated values of the real number's type for maximizing WCE can be real numbers. The number of scenarios $N_{scenarios}$ can be up to $1.25850816 \times 10^{15}$ by Equation (4.6) based on the construction of scenarios. Since the model is stochastic, the number is not finished. The number of simulations $N_{simulations}$ is actually run 6.2925408×10^{15} times when being done five times per a scenario by Equation (4.7).

$$N_{scenarios} = \prod_{k=1}^{13} n(X_k) = 1.25850816 \times 10^{15} \tag{4.6}$$

$$N_{simulations} = 5\, N_{scenarios} = 5 \prod_{k=1}^{13} n(X_k) = 6.2925408 \times 10^{15} \tag{4.7}$$

4.2.1.1.2 Validation of the conceptual model

The validation is done by the SMEs' reviews according to processes explained in the methodology chapter. The problem domain is how to generate a MOE and MOPs to estimate WCEE. The review scope is to evaluate appropriateness of the potential MOE and MOPs, input and output data, and scenarios in the conceptual model. The assessment method is reviews using the five-point Likert Scale by SMEs as the informal testing. The assessment criteria mean the assessment-level thresholds which are the potential MOE of three points, the potential MOPs of three points, the input data of three and half points, the output data of three and half points, and the scenarios of three points. Three SMEs are evaluated and selected. They have a background for the subject of WCE analytics and defense. The review results are as shown in Figure 28.

Since all the assessment-levels are larger than each requirement-level, the conceptual model is accepted.

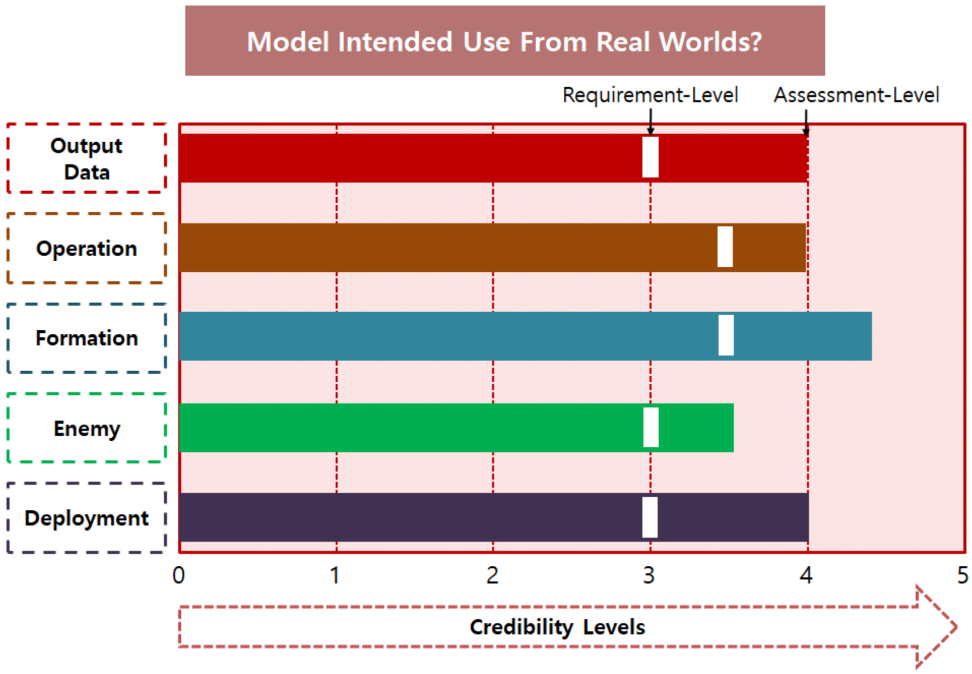

Figure 28. Review Results by SMEs on Conceptual Model Validation

<u>4.2.1.2 A computational model and simulation</u>

The computational model is constructed and implemented from the input data of the conceptual model. The computational model is built by combining two VC simulation models using DIS and HLA. Various scenarios are simulated based on the constructed model. The computational model is verified and validated for the model's reliability.

4.2.1.2.1 A computational model construction for VC simulations

A computational model is constructed with the VC simulation models which are the SIMbox and VR-Forces. The SIMbox and VR-Forces are developed using C++ programming language. SIMbox can play roles as both of virtual and constructive simulation models whereas VR-Forces can play roles as a constructive simulation model. Each simulation model is operated on desktop computers, specifications of which are a Central Processing Unit (CPU) of Inter(R) Core^{TM} i7-4770K, a Random-Access Memory (RAM) of 16GB, a Hard Disk Drive (HDD) of 1TB, a Video Graphic Array (VGA) of NVIDIA GeFrce GTX 770, and a Window 7 Operating System (OS). The SIMbox Development Toolkit is used to model agents that are simulated on KnowBook for a virtual or constructive simulation. MÄK VR-Forces is used for a constructive simulation and MÄK Data Logger is operated for simulation recording and replay.

Distributed in different spots, the heterogeneous VC simulation models are connected to each other through MÄK RTI in order to build a federation which is the computational model. The federation is composed of three federates: a) VR-Forces Simulation Engine (Federate 1), b) VR-Forces GUI (Federate 2), and c) SIMbox (Federate 3). The federation uses the rtiexec approach for RTI developers to manage centralized knowledge. Figure 29 shows a computational model construction for VC simulations.

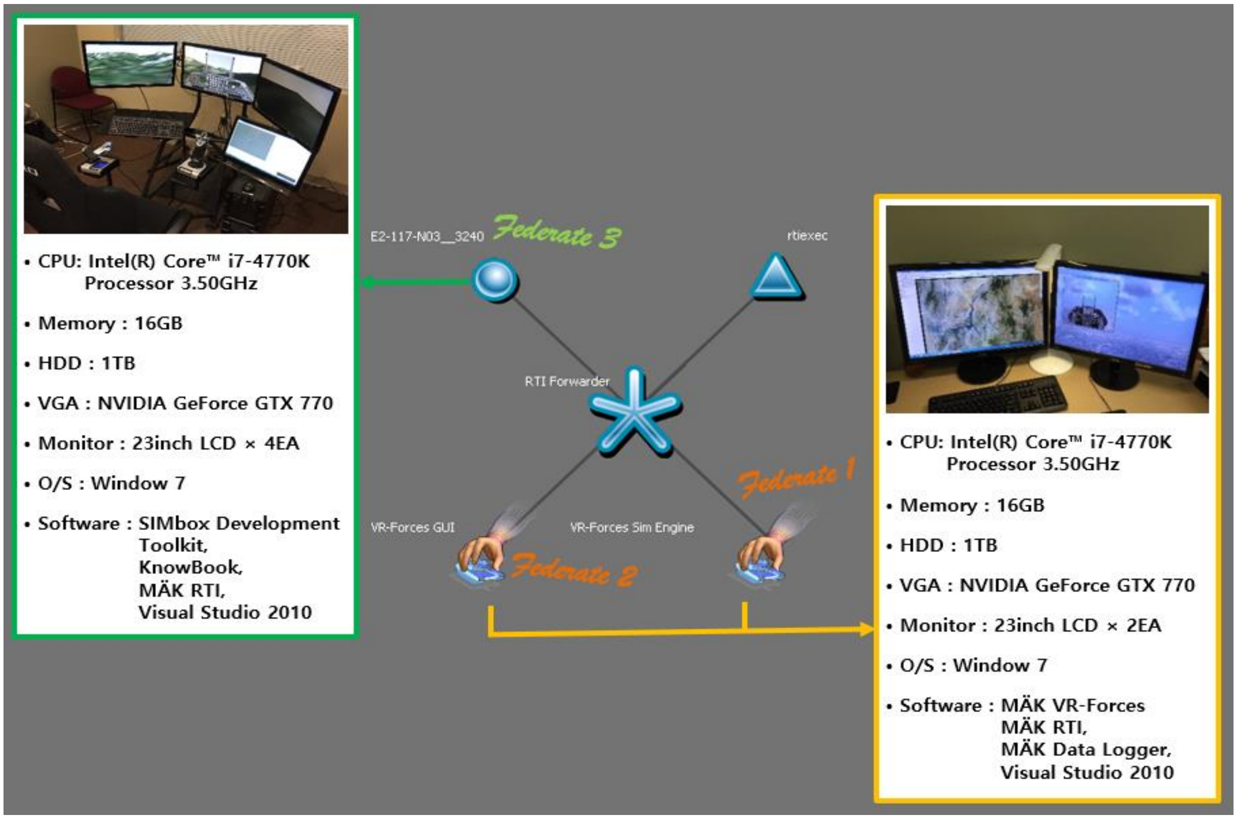

Figure 29. Computational Model Construction for VC simulations

4.2.1.2.2 Implementation of input data into the computational model

Input data based on scenarios designed in the conceptual model is implemented to the computational model using Graphical User Interface (GUI). Figure 30 represents a scenario focused example of the computational model and input data GUI used to implement CGFs for VC simulations. The computational model is composed of seven F-16s, two SAMs, a core facility, and a support system to generate data during a scenario execution. An F-16 is an agent for virtual simulations whereas F-16s, SAMs, a core facility, and a support system are agents for the constructive simulations. The F-16 and SAM have the separate sub-agents which are a M-117 and a SA-8 other than their components.

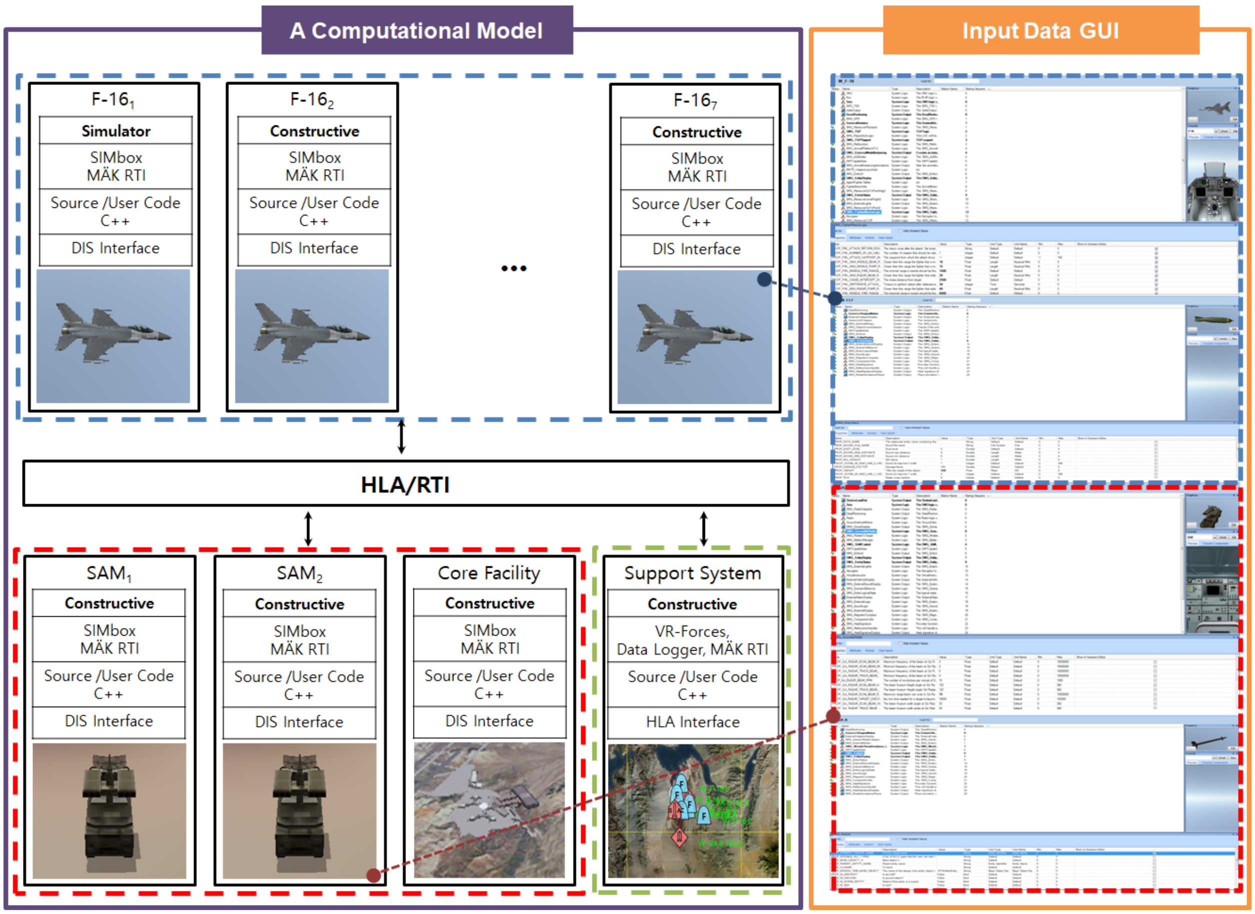

Figure 30. Scenario Focused Example of a Computational Model and Input Data GUI

The input data is implemented to model five types of agents which are a F-16, a SAM, a SA-8 missile, a M-117, and a core facility. They are implemented into the computational model using GUI. Table 14 shows the values on unfixed input data of F-16s and SAMs in the scenario.

Table 14. Example of Values on Interesting Types of Input Data of F-16s and SAMs

Factors ($X_{\bullet j}$)	F-16						SAM								
	X_1	X_2	X_8	X_{10a}	X_{10b}	X_{11}	X_3	X_4	X_5	X_6	X_7	X_9	X_{12a}	X_{12b}	X_{13}
Value	5	40	10	0	0	7	10	20	10	20	6	20	1	0	2

Figure 31 shows the agents' deployment from the initial setting implemented by input data to simulate the above scenario. The agents are deployed in Las Vegas area which is 160 × 70 miles^2. Weather is sunny and temperature reaches to 82 degrees Fahrenheit in the area. Table 15 represents initial statuses of each F-16 and SAM.

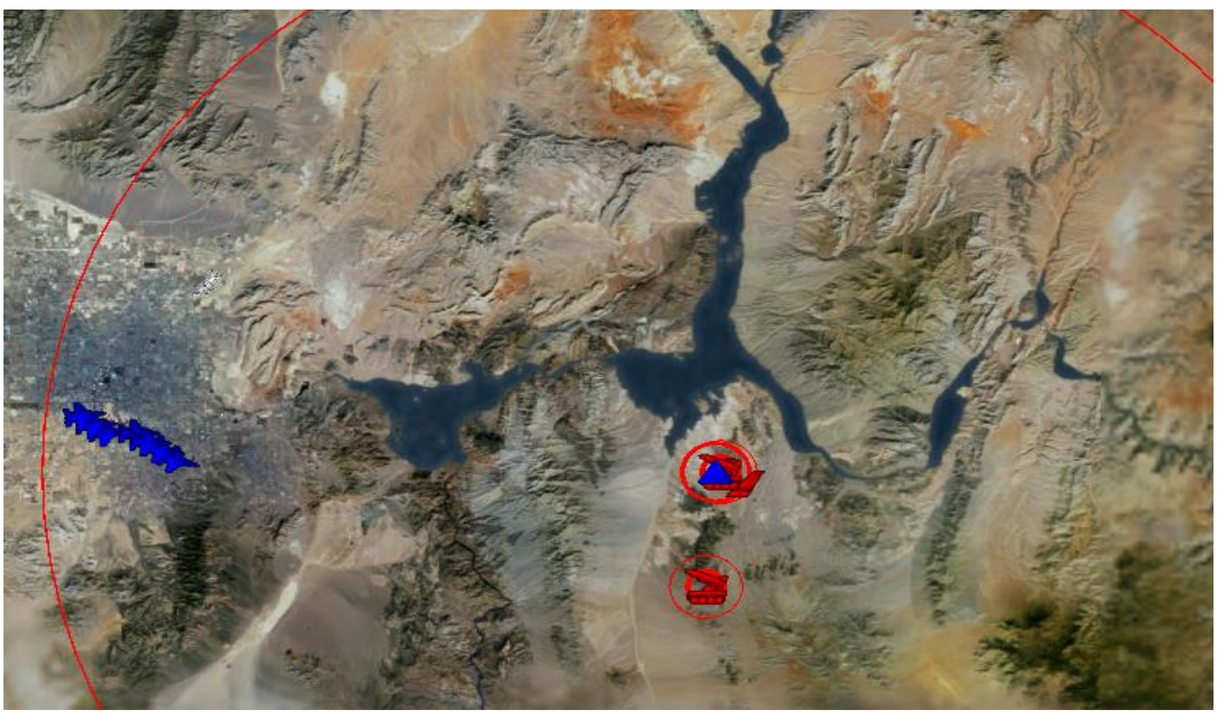

Figure 31. Initial Agents' Deployment Implemented by Input Data

Table 15. Initial Status Example on F-16s and SAMs

Status Factors	F-16							SAM	
	1	2	3	4	5	6	7	1	2
AGL(ft)	7768	7696	7631	7710	7738	7698	7713	0	0
Mach	0.51	0.53	0.51	0.52	0.51	0.53	0.52	0	0
G (°)	-2.69	-2.65	-2.73	-2.68	-2.61	-2.67	-2.65	0	0
Heading(°)	92	88	94	90	90	89	92	270	282
Latitude(°)	N36: 02.12	N36: 02.43	N36: 03.16	N36: 03.53	N36: 04.13	N36: 04.30	N36: 05.07	N36: 01.22	N36: 53.31
Longitude(°)	W115: 04.18	W115: 05.19	W115: 06.17	W115: 07.07	W115: 08.57	W115: 10.18	W115: 11.08	W114: 24.22	W114: 25.29

As discussed in 4.1, the Overview Section, one of this case study's purposes is to develop a pilot model to show processes of the suggested methodology on WCE analytics. Also, an experiment environment has limitations; the number of simulators and the number of experiment participants due to restricted budget of Simulation Interoperability Laboratory at University of Central Florida. Therefore, virtual simulations are operated by Artificial Intelligence (AI). Input data for the observation error and the reaction time automatically reflected by weapon operators is manually substituted with the damage factor and the armor factor. The factors are coded to generate and analyze data on the observation error and reaction time in a virtual simulation.

4.2.1.2.3 Simulation

The computational model implemented by the conceptual model should be simulated based on the number of scenarios calculated by Equation (4.5). However, the experiment environments have limitations on budget and time. So, samples of three hundred scenarios in this case study are uniformly chosen to solve this problem. At this time, the samples should cover all attributes of potential MOPs to estimate a relationship between a MOE and all MOPs. Each scenario is done five times and, thus, the number of simulations reaches to 1,500. Figure 32 shows a simulation example through the computational model. The output data for the structured data is generated by the computational model as shown in Equation (4.8).

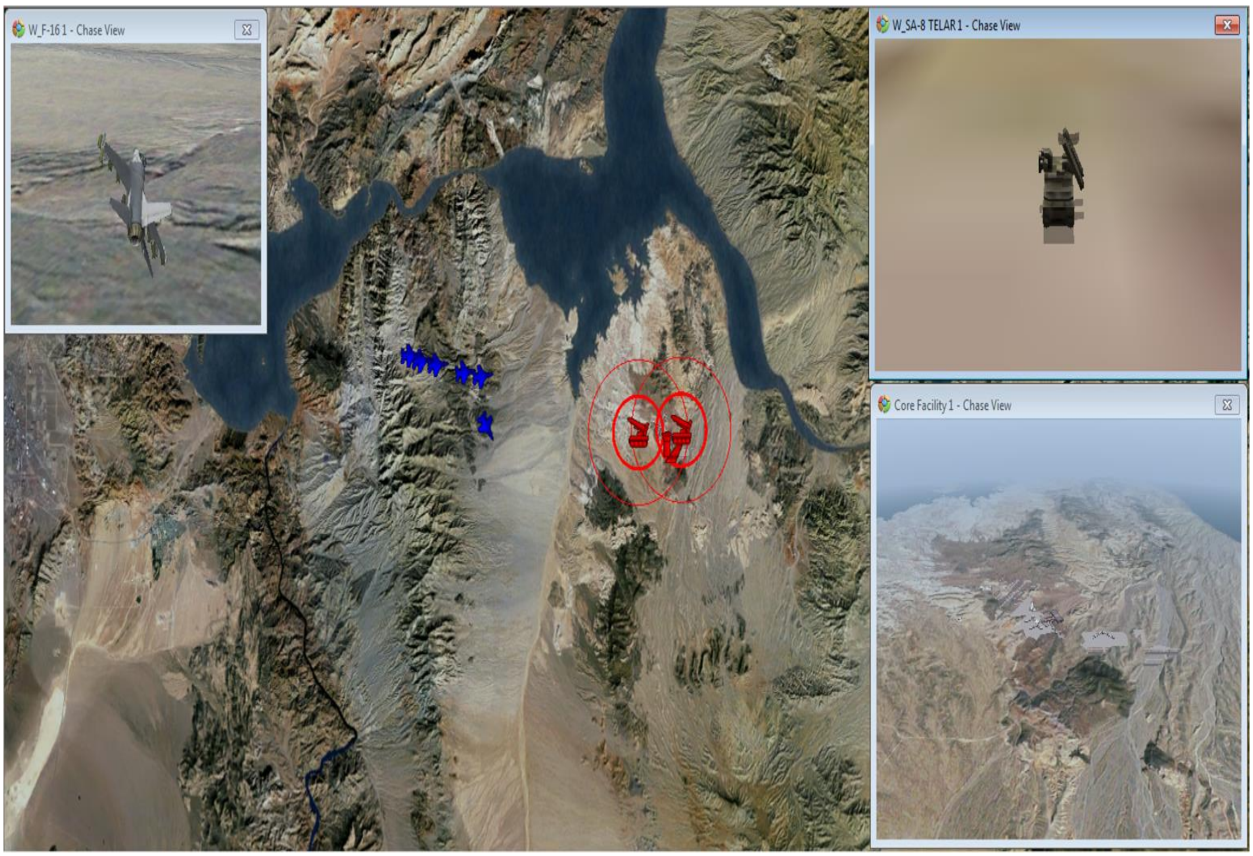

Figure 32. Simulation Example

δ_{CM}: $\boldsymbol{ID} \rightarrow \boldsymbol{OD_s}$ =

$$\left\{ O_{\,i,l,m,n} \,\middle|\, O_{i,l,m,n} = \begin{pmatrix} \boxed{\begin{matrix} 7 & 0 & \cdots & \text{Red} \\ 39 & 0 & & \text{Blue} \\ \vdots & & \ddots & \vdots \\ 45 & 0 & \cdots & \text{Red} \end{matrix}} & , \cdots , & \boxed{\begin{matrix} 7 & 0 & \cdots & \text{Red} \\ 39 & 0 & & \text{Blue} \\ \vdots & & \ddots & \vdots \\ 45 & 100 & \cdots & \text{Red} \end{matrix}} \\ \boxed{\begin{matrix} 5 & 0 & \cdots & \text{Red} \\ 9 & 0 & & \text{Blue} \\ \vdots & & \ddots & \vdots \\ 13 & 0 & \cdots & \text{Blue} \end{matrix}} & , \cdots , & \boxed{\begin{matrix} 5 & 100 & \cdots & \text{Red} \\ 9 & 100 & & \text{Blue} \\ \vdots & & \ddots & \vdots \\ 13 & 100 & \cdots & \text{Blue} \end{matrix}} \\ & \vdots & \\ \boxed{\begin{matrix} 15 & 0 & \cdots & \text{Blue} \\ 3 & 0 & & \text{Red} \\ \vdots & & \ddots & \vdots \\ 12 & 0 & \cdots & \text{Blue} \end{matrix}} & , \cdots , & \boxed{\begin{matrix} 15 & 100 & \cdots & \text{Blue} \\ 3 & 100 & & \text{Red} \\ \vdots & & \ddots & \vdots \\ 12 & 100 & \cdots & \text{Blue} \end{matrix}} \end{pmatrix} \right\} \tag{4.8}$$

δ_{CM}: Computational Model of the WCE

$\boldsymbol{OD_s}$: Output Data Set for the Structured Data

$O_{i,l,m,n}$: Value on i_{th} Simulation, l_{sec} Time, m_{th} Agent, and n_{th} Attribute

4.2.1.2.4 Verification and validation on the computational model

The computational model, which uses default values offered by the SIMBox toolkit, has already been verified and validated. This is because the SIMbox toolkit is developed based on the strict quality management system. For example, an F-16 agent in the SIMbox is tested and developed in about 2 years by engineers through the verification and validation processes. Also, the ISO 9001:2008 certification processes are applied as a core part of the SIMbox quality management system. It can guarantee the SIMbox's high quality and continuous improvement in terms of its verification and validation (http://www.simigon.com/quality-policy.html).

Apart from the SIMbox's quality management system, the computational model is also verified and validated with other ways. The computational model is verified on whether the input data is implemented into computational model correctly by using an animation approach. The method is to monitor the animation results for the input data implementation to the computational model. For example, Figure 33 shows an animation for the verification of the SAM's missile range and type. Implementation of the SAM's missile range and type is monitored to tell if the SAM launches the SA-8s according to the input data by using the chase view. Also, the number of and type of bombs loaded and launched in F-16 are verified as shown in Figure 34. The computational model is verified by the animation monitoring as shown in the above examples.

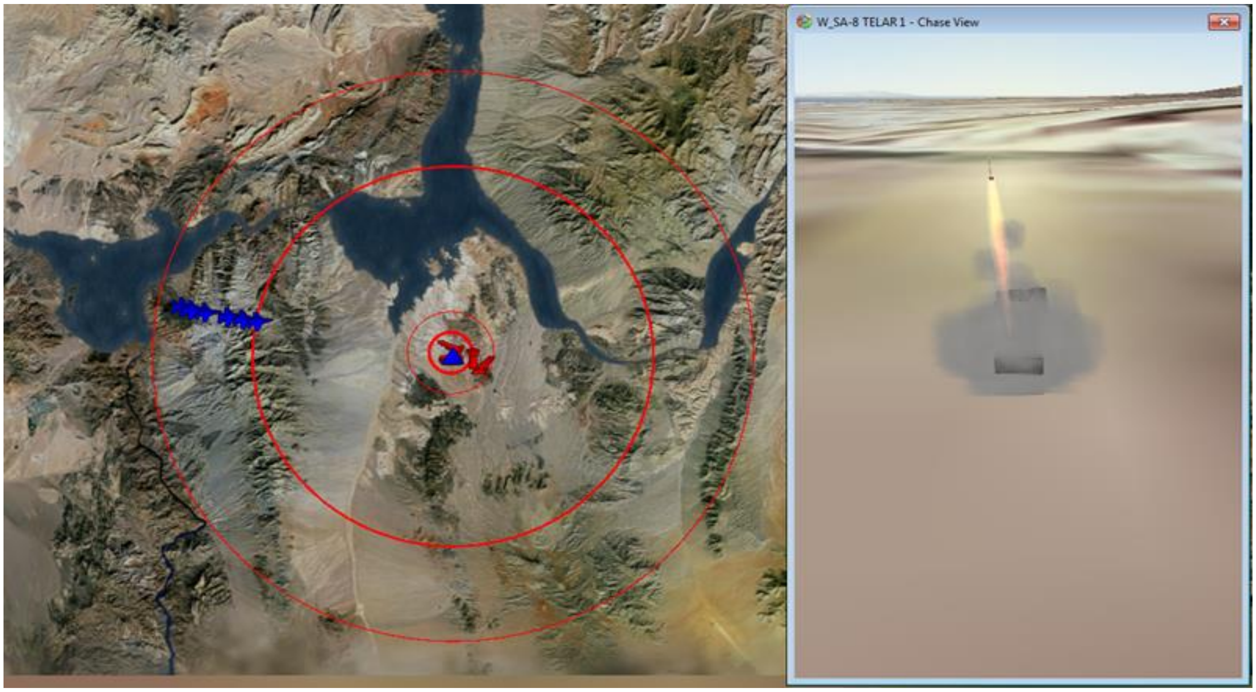

Figure 33. Verification on SAM's Missile Range and Type by Animation (SIMbox)

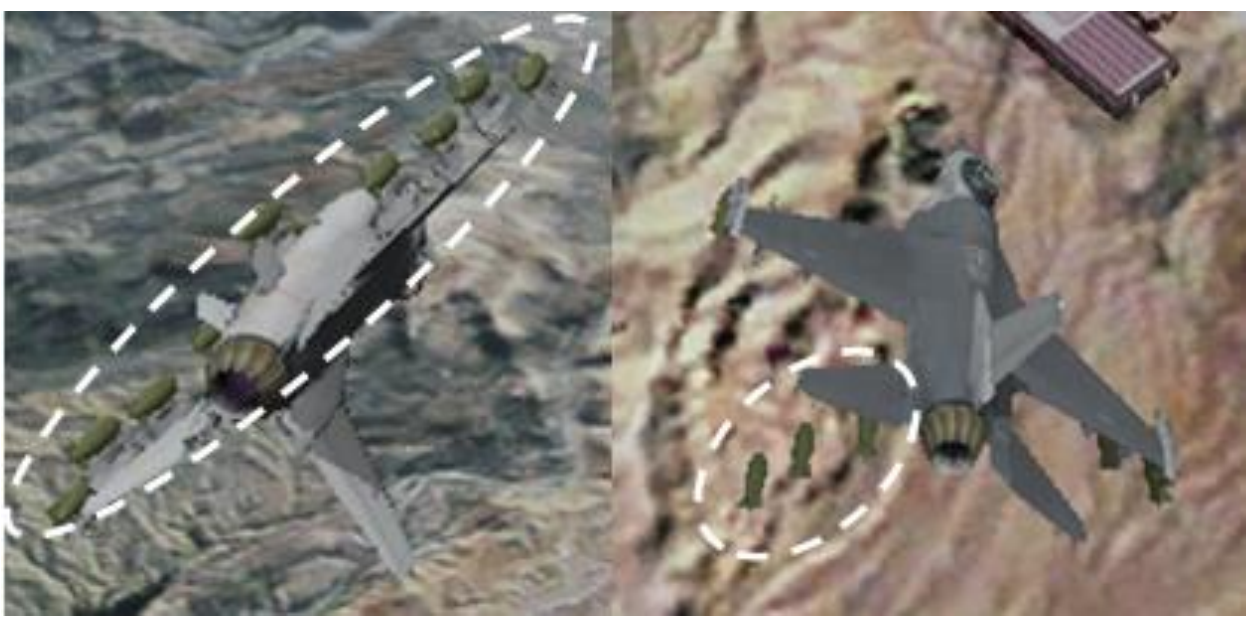

Figure 34. Verification on the Number of and a Type of Bombs Loaded and Launched by Animation (SIMbox)

The validation activity is seamlessly done during simulations. Figure 35 shows an example of the validation activity done on whether the SA-8s are launched and attack F-16s with the correct operation format compared to the real world. The computational model is also validated by the SMEs' review other than by the SIMbox Toolkit quality management system. The problem domain is to validate as to whether the computational model is the right model reflecting the real world for WCE analytics. The review scope is to evaluate pertinence of output data and operational factors simulated by the computational model compared to the real world. The operational factors are considered to be the distributed operation, the operationally durable formation, the distributed enemy, and the distributed deployment. The SMEs review the simulation results by using the five-point Likert Scale as the informal testing.

The assessment criteria are the output data of three points, the distributed operation of three and half points, the operationally durable formation of three and half points, the distributed enemy of three points, and the distributed deployment of three and half points as the requirement-level thresholds. Three SMEs, who have a background on the subject of WCE analytics and defence, are evaluated and selected. Figure 36 shows the results reviewed by the SMEs. All the assessment-levels are larger than each requirement-level, and thus, the computational model is validated.

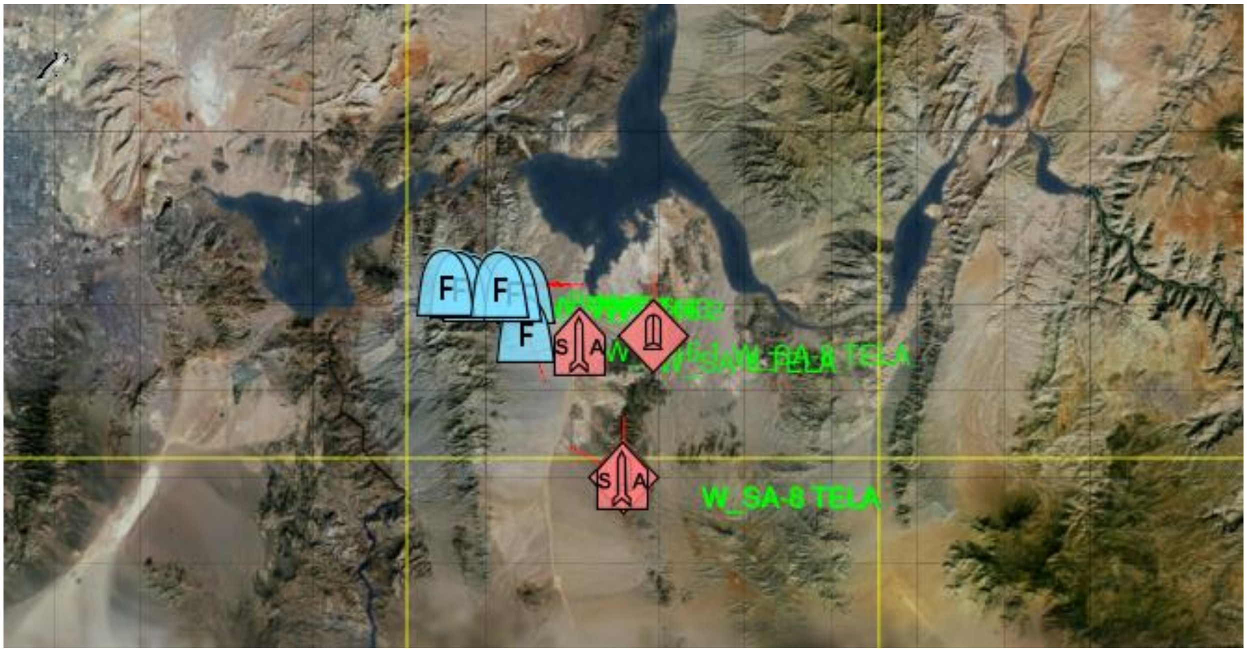

Figure 35. Validation on SA-8s launch and attack against F-16s by Animation (VR-Forces)

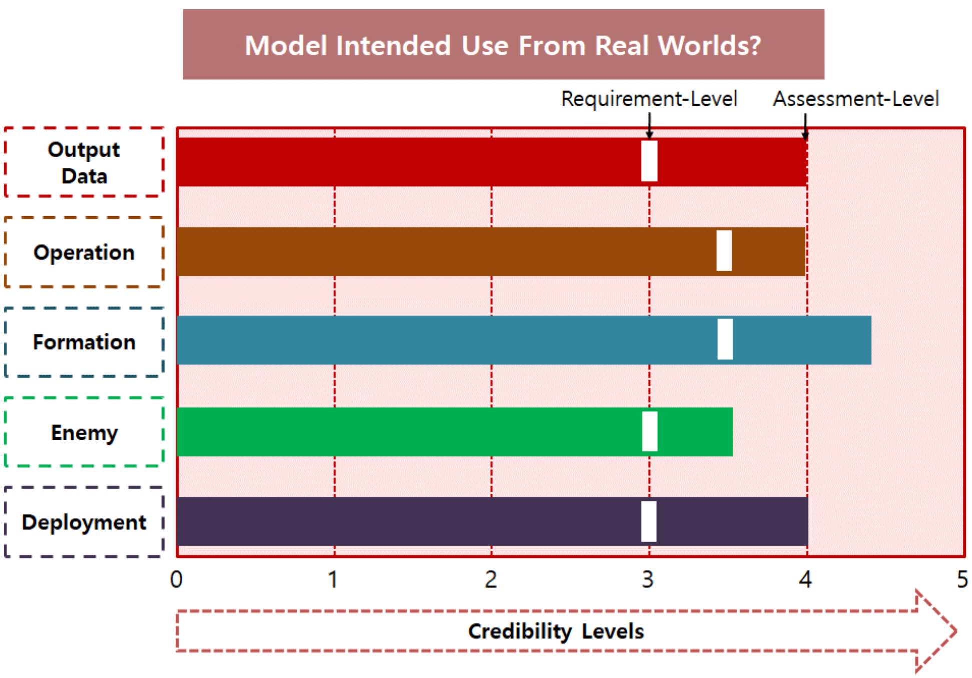

Figure 36. Review Results by SMEs on Computational Model Validation

4.2.2 STEP 2: Collecting Additional Data

The additional data related to CE can be collected by various sources. This case study applies the human input data but not existing models and other sources. Also, the additional data set for MOPs are not considered. The constraints of the human input data are reflected and gathered by the data engineer system and the stakeholder system. The constraints' range has to be set by the stakeholder system's rational decision based on objective facts. The collected data is validated by the SMEs. Lower and upper bounds of the constraints applied to this case study are assumed as shown in Equation (4.9) and Equation (4.10).

$\boldsymbol{\theta}_{SS}$, $\boldsymbol{\xi}_{DS} \rightarrow \boldsymbol{AD_b}=$

$$\left\{ad_{b_j} \mid a_j \le ad_{b_j} \le b_j\right\} \tag{4.9}$$

$$\rightarrow \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{A}_j \\ \boldsymbol{B}_j \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 11 & 13 \\ 1 & 6 & 5 & 10 & 0 & 5 & 3 & 1 & 10 & 1 & 1 \\ 5 & 10 & 20 & 20 & 19 & 20 & 5 & 10 & 50 & 3 & 2 \end{bmatrix} \tag{4.10}$$

$\boldsymbol{\theta}_{SS}$: Stakeholder System

$\boldsymbol{\xi}_{DS}$: Data Engineer System

$\boldsymbol{AD_b}$: Additional Data Set for Constraints

ad_{b_j}: Additional Data Value for Constraints on j_{th} Attribute

$\boldsymbol{J} = \{1, 2, \ldots, 9, 11, 13\}$: Constraints Index

$\boldsymbol{A}_j = \{a_1, a_2, \ldots, a_9, a_{11}, a_{13}\}$: Lower Bounds

$\boldsymbol{B}_j = \{b_1, b_2, \ldots, b_9, b_{11}, b_{13}\}$: Upper Bounds

$\mathbb{Z} \ni ad_{b_1}, ad_{b_2}, ad_{b_7}, ad_{b_{11}}, ad_{b_{13}}$

$\mathbb{R} \ni ad_{b_3}, ad_{b_4}, ad_{b_5}, ad_{b_6}, ad_{b_8}, ad_{b_9}$

The acquisition cost of the weapon systems can be a limitation to maximizing the WCE. Equation (4.11) shows a constraint on the acquisition cost of the F-16 and the M-117. The cost constraint is added to the extracted constraints above. Some of these constraints, which are assumed in this case study, are used to find the optimal values to maximize the MOE based on the WCEE using MIP.

$\boldsymbol{\theta}_{SS}, \boldsymbol{\xi}_{DS} \rightarrow \boldsymbol{AD}_{b_{14}} =$

$$\{(ad_{b_1}, ad_{b_{11}}) | 18.8ad_{b_{11}} + 0.17ad_{b_1} \leq 52\} \tag{4.11}$$

$\boldsymbol{AD}_{b_{14}}$: Additional Data Set for 14_{th} Constraints

4.2.3 STEP 3: Processing Big Data for F-16 CE Analytics

The generated and collected big data are processed for F-16 CE analytics. The data is selected from the big data by the expert system, and then it is cleaned, integrated, and transformed by using the data preparation process. After the processes the data is stored into HDFS and then extracted by using MapReduce process to ready for F-16 analytics. The verification and validation are done while processing the big data.

4.2.3.1 Selection

The output data generated by the VC simulations has various subset matrixes, one of which is selected to estimate the WCEE as shown in Equation (4.12). All the human input data in an additional data collection step is chosen by the expert system. The SMEs' reviews are used to validate the data selected by the expert system. The selected data is checked using a five-point Likert Scale by the SMEs. The assessment-level of four points is larger than the requirement-level of three points. Therefore, the selected data is accepted based on the testing result.

$\boldsymbol{\theta_{ES}}$: $\boldsymbol{OD_s}$, $\boldsymbol{AD_a}$, $\boldsymbol{AD_b} \rightarrow \boldsymbol{SD}$ =

$$\left\{ sd_{p,q,r,s} \;\middle|\; sd_{p,q,r,s} = \begin{pmatrix} \begin{matrix} 0 & 0 & \cdots & 50 \\ 0 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 200 & \cdots & 20 \end{matrix} & , \cdots , & \begin{matrix} 38 & 0 & \cdots & 50 \\ 100 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 200 & \cdots & 20 \end{matrix} \\ \begin{matrix} 0 & 0 & \cdots & 50 \\ 0 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 200 & \cdots & 20 \end{matrix} & , \cdots , & \begin{matrix} 100 & 0 & \cdots & 50 \\ 72 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 200 & \cdots & 50 \end{matrix} \\ & \vdots & \\ \begin{matrix} 0 & 2000 & \cdots & 10 \\ 0 & 2000 & & 10 \\ \vdots & & \ddots & \vdots \\ 0 & 2000 & \cdots & 50 \end{matrix} & , \cdots , & \begin{matrix} 72 & 2000 & \cdots & 10 \\ 0 & 2000 & & 10 \\ \vdots & & \ddots & \vdots \\ 0 & 2000 & \cdots & 50 \end{matrix} \end{pmatrix} \right\} \tag{4.12}$$

$\boldsymbol{\theta_{ES}}$: Expert System

$\boldsymbol{OD_s}$: Output Data Set for the Structured Data

$\boldsymbol{AD_a}$: Additional Data Set for MOPs

$\boldsymbol{AD_b}$: Additional Data Set for Constraints

$\boldsymbol{SD}$: Selected Data Set

$sd_{p,q,r,s}$: Value on p_{th} Simulation, q_{th} Time, r_{th} Agent, and s_{th} Attribute

The amount of selected data per simulation can be various according to the number of agents and considered attributes as well as the simulation time. If 67 agents including 95 considered attributes are simulated during 350 seconds 1,500 times, the selected data reaches to 1.13 TB.

4.2.3.2 Cleaning, integration, and transformation

The data generated by operators should be cleaned because the human data generated by the interaction between an operator and a computational model usually has much noise and inconsistent characteristics. The data generated by the VC simulations is integrated with the data attained from operators' actions. Since the simulators are executed by AI in this case study, these steps are omitted. The integrated data is transformed into a proper form for the WCE analytics. The observation error and reaction time measured during the F-16 and SAM simulators' operation is normalized to index which has values of 1 through 100.

4.2.3.3 Storage and extraction

The total data is constructed from the selected data through the data preparation process. It is stored into HDFS built on an Ubuntu 17.10 operating system by using a Java Virtual Machine (JVM). Also, the reduced data is extracted from the total data through the MapReduce process for data analytics. The total data and reduced data are validated and verified through the SMEs' reviews and debugging the code.

4.2.3.3.1 Storage to HDFS

The data integrated and transformed is stored to the HDFS which are constructed on an Ubuntu 17.10 operating system. In this case study the HDFS architecture is composed of a user node, a name node, a secondary name node, and three data nodes. The user node is an interface between a user and the HDFS to deal with data. The name node has all the HDFS metadata which can be a file name, the number of the block replication, location of blocks on three data nodes, file attributes data, etc. Data storage is done using a JVM by a user in a user node. The secondary name node keeps information about the metadata of a name node for insurance. The three data nodes have file contents which are separated to blocks. The big data storage process has eight steps as shown in the following: a) access to Distributed File System (DFS), b) create a file which has the metadata kept in RAM, c) order data storage using the file system data output stream, for which d) receive the metadata from a name node, e) split a file that has 1,500 simulations' data to blocks, each of which has 64MB, f) store the data to three data nodes as blocks, each of which is replicated three times as a pipeline format, g) report information on blocks stored to the name node, and h) close a task of data storage. The big data storage process on the HDFS architecture is shown in Figure 37. Equation (4.13) presents a set of total data stored into HDFS using the big data storage process. The total data, which is transformed from the selected data by the data preparation process, is validated by the SMEs' reviews. The five-point Likert Scale is used for the SMEs to validate the total data. Since the assessment-level is four points and the requirement-level is three and half points, the total data is accepted.

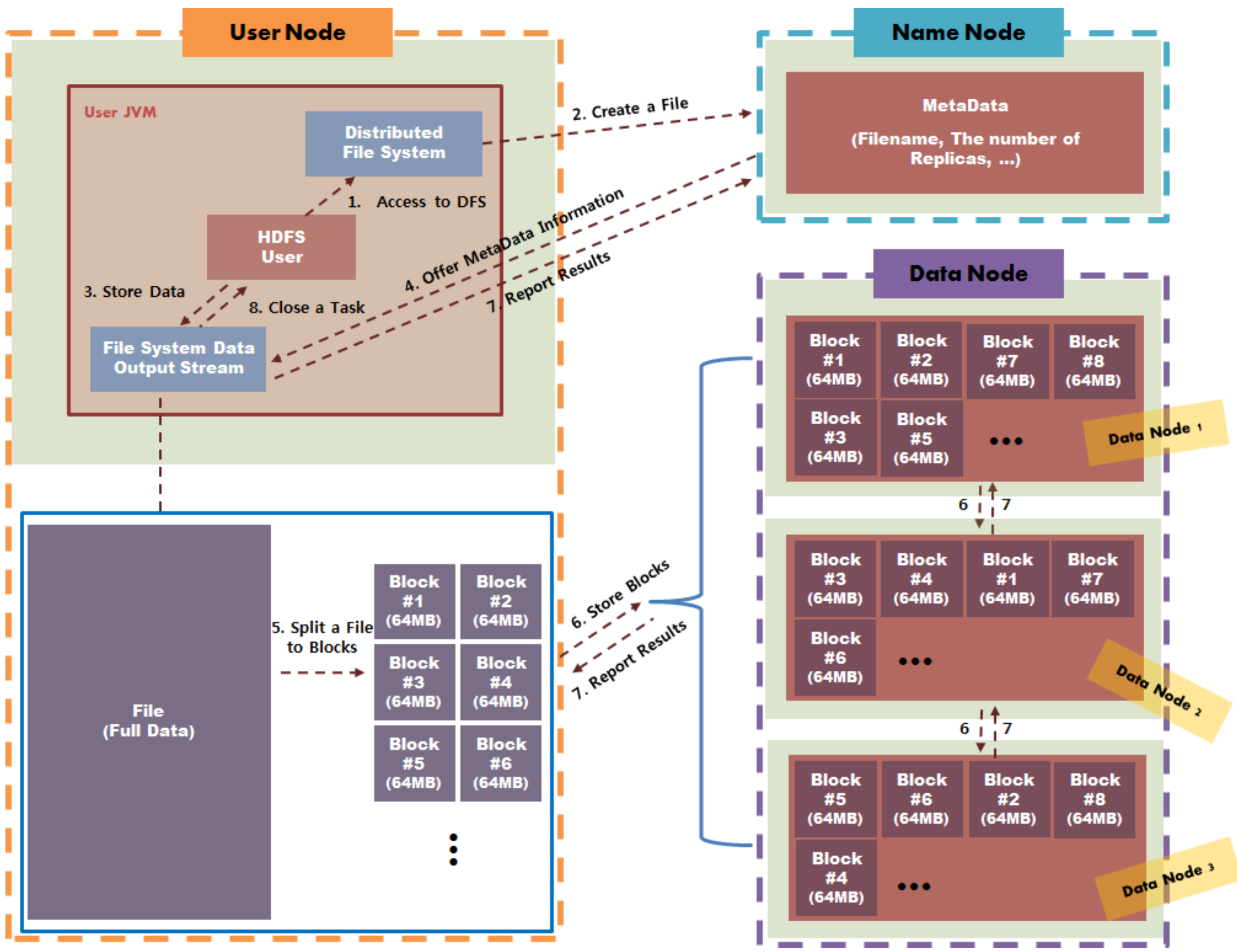

Figure 37. Big Data Storage Process

$\boldsymbol{\psi_{DP}}$: $\boldsymbol{SD \rightarrow TD_s} =$

$$\left\{ td_{p,q,r,s} \;\middle|\; td_{p,q,r,s} = \begin{pmatrix} \begin{matrix} 0 & 0 & \cdots & 50 \\ 0 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 0.2 & \cdots & 20 \end{matrix}, \cdots, & \begin{matrix} 38 & 0 & \cdots & 50 \\ 100 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 0.2 & \cdots & 20 \end{matrix} \\ \begin{matrix} 0 & 0 & \cdots & 50 \\ 0 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 0.2 & \cdots & 20 \end{matrix}, \cdots, & \begin{matrix} 100 & 0 & \cdots & 50 \\ 72 & 0 & & 50 \\ \vdots & & \ddots & \vdots \\ 0 & 0.2 & \cdots & 50 \end{matrix} \\ \vdots & \\ \begin{matrix} 0 & 2 & \cdots & 10 \\ 0 & 2 & & 10 \\ \vdots & & \ddots & \vdots \\ 0 & 2 & \cdots & 50 \end{matrix}, \cdots, & \begin{matrix} 72 & 2 & \cdots & 10 \\ 0 & 2 & & 10 \\ \vdots & & \ddots & \vdots \\ 0 & 2 & \cdots & 50 \end{matrix} \end{pmatrix} \right\} \tag{4.13}$$

$\boldsymbol{\psi_{DP}}$: Data Preparation Process

$\boldsymbol{TD_s}$: Total Data Set for the Structured Data

$td_{p,q,r,s}$: Value on p_{th} Simulation, q_{th} Time, r_{th} Agent, and s_{th} Attribute

4.2.3.3.2 MapReduce process

The reduced data is acquired by the MapReduce process. The Mapper and the Reducer are coded to implement the MapReduce process on the Java computer programming language as shown in Figure 38.

```java
public static class TokenizerMapper     extends Mapper<LongWritable, Text, Text, IntWritable>{

        private final static IntWritable outputValue = new IntWritable(1);
        private Text outputKey = new Text();

        public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
                SimulationParser parser = new SimulationParser(value);
                outputKey.set(parser.getSystemTime() + "," + parser.getUniqueName() + "," + parser.getEntityID() +
                                parser.getDamageValue() + "," + parser.getWhichSide() + "," + parser.getArmorFactor() +
                                parser.getDamageFactor() + "," + parser.getKillRadius() + "," + parser.getIsAircraft() +
                                parser.getIsGround() + "," + parser.getIsWeapon());

                if (parser.getDamageFactor() == 100 && parser.getDamageValue() == 1) {
                        context.write(outputKey, outputValue);
                }
        }
}
```

Mapper

```java
public static class IntSumReducer extends Reducer<Text,IntWritable,Text,IntWritable> {
        private IntWritable result = new IntWritable();

        public void reduce(Text key, Iterable<IntWritable> values,
                        Context context
                        ) throws IOException, InterruptedException {
                int sum = 0;
                for (IntWritable val : values) {
                        sum += val.get();
                }
                result.set(sum);
                context.write(key, result);
        }
}
```

Reducer

Figure 38. Mapper and Reducer Code to Implement MapReduce Process (Java)

The Mapper has two main factors which are the key and the values. The key is that the core facility of the red team is destroyed. The values are the F-16s survival rate and the primary potential MOPs. The Mapper does mapping to extract the values from three data nodes, given that the key is satisfied per simulation.

The Reducer helps to report data results reduced from three data nodes by the Mapper. The 930 of 1,500 simulations, which are stored to three data nodes, satisfy the F-16 mission that is to destroy the core facility of the red team. The 930 simulations become the analytics scope based on the key set in the Mapper as shown in Figure 39.

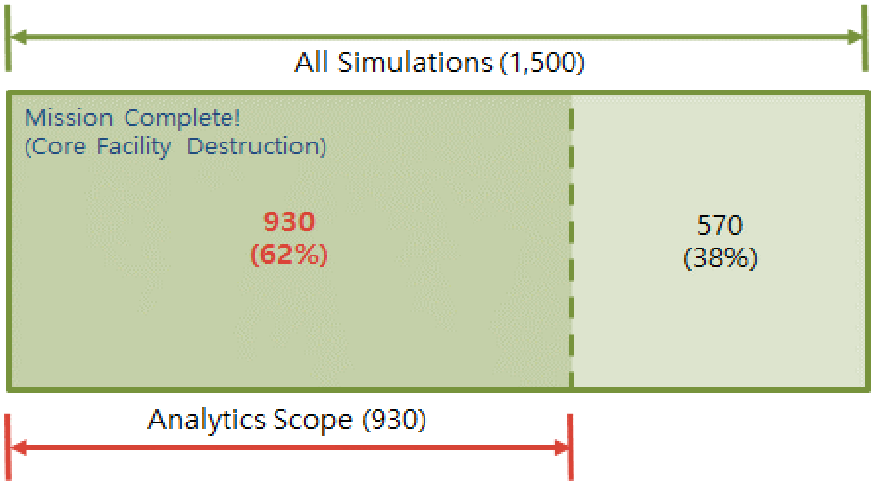

Figure 39. Analytics Scope

4.2.3.3.3 Extraction from HDFS

The reduced data is extracted from the big data stored into the HDFS as shown in the followings: a) open the Distributed File System (DFS), b) request the block locations in the metadata kept in the RAM, c) order the data extraction using the file system data input stream according to the MapReduce process, d) receive the block locations in the metadata from a name node, e) request the blocks to be extracted to three data nodes, f) get the requested blocks from three data nodes, g) integrate all extracted blocks to a file, and h) get the file and close the extraction task. The big data extraction process on the HDFS architecture is shown in Figure 40.

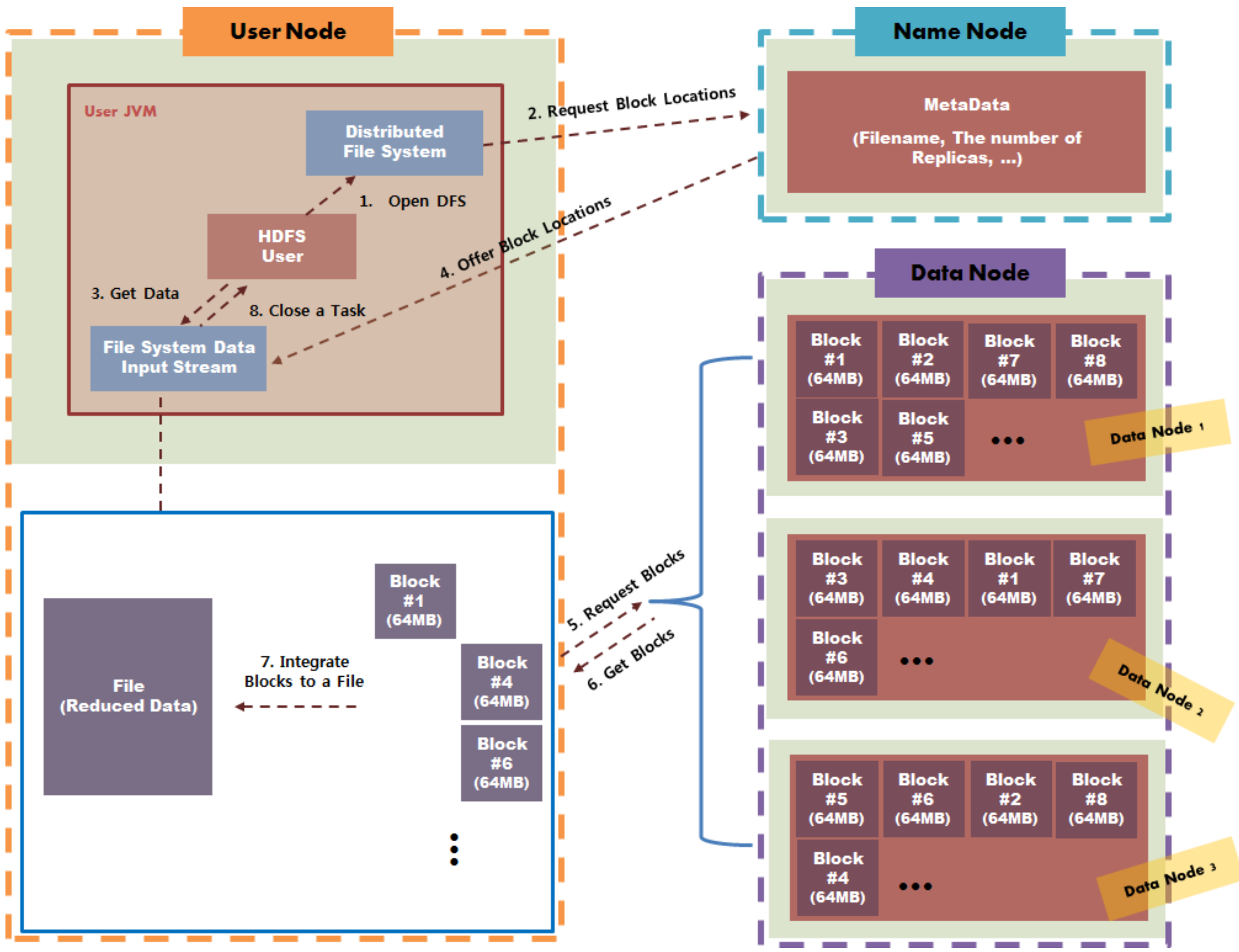

Figure 40. Reduced Data Extraction Process

The reduced data is a subset extracted from the 930 of 1,500 simulations' data set. The subset has 15 attributes as shown in the following: a) F-16s survival rate, given that the core facility of the red team is destroyed , b) number of release bombs, c) number of chaffs and flares, d) control fire range, e) control track range, f) difference between control fire and track range, g) control time between fires, h) number of missiles, i) observation error, j) reaction time, k) dummy attribute Ⅰ for three attack methods, l) dummy attribute Ⅱ for three attack methods, m) number of attack flights, n) dummy attribute Ⅰ for three kinds of deployments, o) dummy attribute Ⅱ for three kinds of deployments, and p) number of SAMs. The first attribute is the primary potential MOE and the others are the primary potential MOPs in the reduced data, consisting of a 930×15 matrix format. Equation (4.14) presents a set of reduced data extracted from the HDFS using the reduced data extraction process.

The reduced data is validated by the SMEs' reviews on whether the transformed data is correct from the total data according to the MapReduced process. The SMEs use a five-point Likert Scale to validate the reduced data. Since the assessment-level of four and half points is more than the requirement-level of three and half points, the reduced data is accepted. The reduced data is verified by debugging the code, when the MapReduce process is implemented to JAVA programming. The reduced data is validated based on the testing result.

$\boldsymbol{\psi_{MP}}$: $\boldsymbol{TD} \rightarrow \boldsymbol{RD}$ =

$$\left\{ rd_{p,q} \middle| rd_{p,q} = \begin{pmatrix} 40 & 6 & \cdots & 1 \\ 40 & 6 & & 1 \\ \vdots & & \ddots & \vdots \\ 83 & 5 & \cdots & 1 \end{pmatrix} \right\} \tag{4.14}$$

$\boldsymbol{\psi_{MR}}$: MapReduce Process

$\boldsymbol{RD}$: Reduced Data Set

$\boldsymbol{TD}$: Total Data Set

$rd_{p,q}$: Value on p_{th} Simulation and q_{th} Attribute (q=1: MOE, q=2,…,q: MOP)

4.3 Analytics Subsystem and Visualization for F-16 CE

This section shows the WCEE development processes and optimal values for maximizing the F-16 CE by the LLM_{AN}. Also, the results are visualized for decision makers to select the best option effectively and efficiently.

4.3.1 STEP 4: Estimating F-16 CE Equation and MOPs' Optimal Values

The F-16 CE equation is estimated by four algorithms for the WCEE development. The MOPs' optimal values are obtained through the equation and constraints to maximize the F-16 CE. Also, the processes and data are verified and validated by the SMEs' reviews, debugging the code, traceability assessment, hypothesis testing, prediction testing, and data analysis methods.

4.3.1.1 F-16 CE equation

Four algorithm pseudo-codes are given to express the processes of the F-16 CE equation development which are the MOPs selection, data attributes analysis, WCEEs development, and WCEE selection. All the pseudo-codes are implemented by R programing language. The F-16 CE equation is evaluated and developed using R programming language.

4.3.1.1.1 MOPs selection

The reduced data is transferred to the normalized data by the normalization process prior to the MOPs selection phase. The expected value of the F-16 survival rate per each scenario is calculated and fifteen MOPs are used per each scenario. The reduced data is normalized as shown in Equation (4.15), Equation (4.16), and Equation (4.17). The normalized data set has 189×16 matrix format.

$\boldsymbol{\psi_{NP}}$: $\boldsymbol{RD} \rightarrow \boldsymbol{ND} = [\boldsymbol{ND_a}\ \ \boldsymbol{ND_b}] =$

$$\left\{ nd_{i,j} \middle| nd_{i,j} = \begin{pmatrix} 40 & 6 & \cdots & 1 \\ 35 & 5 & & 1 \\ \vdots & & \ddots & \vdots \\ 70 & 5 & \cdots & 1 \end{pmatrix} \right\} \tag{4.15}$$

$$\boldsymbol{ND_a} = \{ nd_{i,1} | \ nd_{i,1} = \mathrm{E}(rd_{p,1} | rd_{p,1} \wedge S = i) \} \tag{4.16}$$

$$\boldsymbol{ND_b} = \{ nd_{i,j\neq 1} | \ nd_{i\ ,j\neq 1} = (rd_{p,q\neq 1} | (rd_{p,q\neq 1} \wedge S = i)) \} \tag{4.17}$$

$\boldsymbol{\psi_{NP}}$: Normalization Process

ND: Normalized Data Set

RD: Reduced Data Set

$nd_{i,j}$: Value on i_{th} Scenario and j_{th} Attribute (j=1: MOE, j=2,…,j: MOP)

$rd_{k,p,q}$: Value on p_{th} Simulation and q_{th} Attribute (q=1: MOE, q=2,…,q: MOP)

S: Scenario index (i = 1, … , 189)

The normalized data is used as input data in the MOPs selection phase. The MOPs are chosen by the algorithm pseudo-code for the MOPs selection as shown in Figure 41. The MOPs are selected by the AIC which is a penalized-likelihood criterion. The both mode of stepwise search is applied to the MOPs selection. The MOPs selection is done through six steps which are a process to find the MOPs combination to minimize the AIC. Table 16 shows a final MOPs selection step. Selected are the MOPs which are X1, X2, X3, X7, X8, X9, X10b, X11, X12a, and X12b, because the combination has the minimum AIC.

function MOPs-SELECTION (*ND*) **returns** *WD*

inputs: *ND*, a normalized reduced data set

do SELECTION_VARIABLE ($nd_{i,j \neq 0}$) = *'TRUE'* **then**

y_p, elements of the potential MOE data set $Y \leftarrow nd_{i,0}$

$x_{p,q}$, elements of the potential MOP data set $X \leftarrow nd_{i,j \neq 0}$

returns *WD*, a WCEE data set $\leftarrow \{(y_0, x_{00}, x_{01}, \ldots, x_{0j}), \ldots, (y_p, x_{p,0} \ldots, x_{p,q})\}$

Figure 41. Algorithm Pseudo-Code for MOPs Selection

Table 16. Final MOPs Selection Step by AIC

(Y ~ X1 + X2 + X3 + X7 + X8 + X9 + X10b + X11 + X12a + X12b)

MOPs		DF	Sum of Sq.	RSS	AIC
<none>		-	-	22,298	923.62
-	X9	1	286.8	22,584	924.04
+	X5	1	128.5	22,169	924.53
+	X4	1	128.5	22,169	924.53
-	X8	1	427.4	22,725	925.21
+	X10a	1	17.9	22,280	925.47
+	X6	1	3.6	22,294	925.59
-	X1	1	485.9	22,783	925.7
-	X3	1	733.6	23,031	927.74
-	X12a	1	773.6	23,071	928.07
-	X12b	1	953.5	23,251	929.54
-	X10b	1	1022.1	23,320	930.09
-	X2	1	2,238.1	24,536	939.7
-	X11	1	8,048	30,346	979.87
-	X7	1	12,133.5	3364,431	1,003.74

The linear model as a WCEE is appropriate based on the relationship between the residuals and fitted values by the selected MOPs as shown in Figure 42. Also, the trend of the standardized residuals on the theoretical quantiles is close to linear. This means that the error follows the normal distribution as shown in Figure 43.

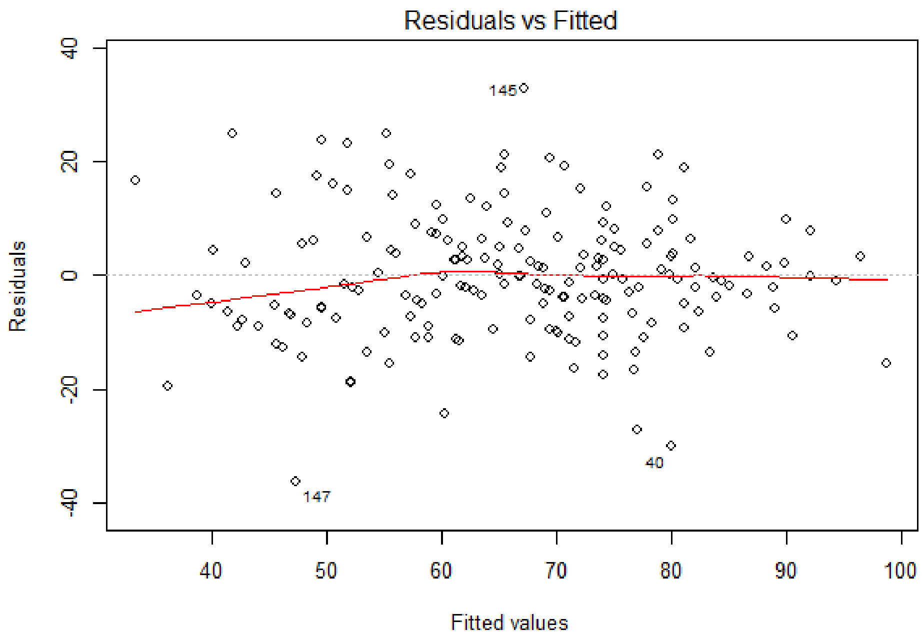

Figure 42. Relationship between Residuals and Fitted Values

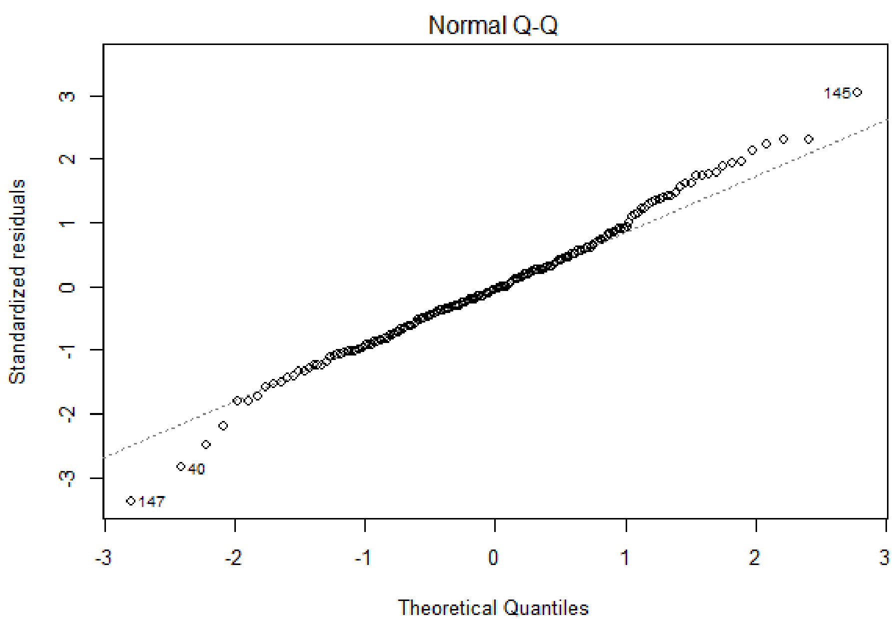

Figure 43. Relationship between Standardized Residuals and Theoretical Quantiles

4.3.1.1.2 Data attributes analysis

Attributes of a data set including the MOE and the selected MOPs are analyzed on whether multicollinearity, outlier, and heteroscedasticity exist or not according to the algorithm pseudo-code as shown in Figure 44. The multicollinearity of the MOPs is analyzed by the VIF which is the proportion of variance in a model with various terms. The VIFs are calculated to identify whether the MOPs have multicollinearity or not as shown in Table 17. Since all the VIFs are under 10, the multicollinearity does not exist in this case. Figure 45 shows the scatterplot matrices which represent the relationship between two MOPs. The multicollinearity existence or not is confirmed by the scatterplot matrices.

function DATA-ATTRIBUTE-ANALYSIS (*WD*) **returns** *WCEE_Array* [ ]

  inputs: *WD*, a WCEE data set

  for p, q = 0 **to** P, Q **do**

    for each $x_{p,q}$ of *WD* **do**

      if MULTICOLLINEARITY? ($x_{\bullet,q}$) = *'TRUE'* **then**

        delete $x_{\bullet,q}$ in *WD* **then**

          if OUTLIER? ($x_{p,\bullet}$) = *'TRUE'* **then**

            delete $x_{p,\bullet}$ in *WD* **then**

              if HATEROSCEDASTICITY? ($x_{p,q}$) = *'TRUE'* **then**

                WCEE_Array [ ] ←

                  \ WEIGHT_REGRESSION (*WD*)

                WCEE_Array [ ] ←

                  \ LOGLINEAR_REGRESSION (*WD*)

              else

```
                                        WCEE_Array [ ] ←
                                            \ MULTIPLE_REGRESSION (WD)
                                if HATEROSCEDASTICITY? (x_{p,q}) = 'TRUE' then
                                    discard this process to develop a WCEE
                                else
                                    WCEE_Array [ ] ← ROBUST_REGRESSION (WD)
                            else
                                if HATEROSCEDASTICITY? (x_{p,q}) = 'TRUE' then
                                    WCEE_Array [ ] ← WEIGHT_REGRESSION (WD)
                                    WCEE_Array [ ] ← LOGLINEAR_REGRESSION (WD)
                                else
                                    WCEE_Array [ ] ← MULTIPLE_REGRESSION (WD)
                        if OUTLIER? (x_{p,•}) = 'TRUE' then
                            delete x_{p,•} in WD then
                            if HATEROSCEDASTICITY? (x_{p,q}) = 'TRUE' then
                                discard this process to develop a WCEE
                            else
                                WCEE_Array [ ] ←
                                    \ PRINCIPAL_COMPONENT_REGRESSION (WD)
                                WCEE_Array [ ] ← RIDGE_REGRESSION (WD)
                                WCEE_Array [ ] ← LASSO_REGRESSION (WD)
                        else
                            if HATEROSCEDASTICITY? (x_{•,q}) = 'TRUE' then
                                discard this process to develop a WCEE
                            else
                                WCEE_Array [ ] ←
                                    \ PRINCIPAL_COMPONENT_REGRESSION (WD)
```

```
                        WCEE_Array [ ] ← RIDGE_REGRESSION (WD)
                        WCEE_Array [ ] ← LASSO_REGRESSION (WD)
        else
            if OUTLIER? (x_{p,•}) = 'TRUE' then
                delete x_{p,•} in WD then
                    if HATEROSCEDASTICITY? (x_{p,q}) = 'TRUE' then
                        WCEE_Array [ ] ← WEIGHT_REGRESSION (WD)
                        WCEE_Array [ ] ← LOGLINEAR_REGRESSION (WD)
                    else
                        WCEE_Array [ ] ← MULTIPLE_REGRESSION (WD)
                if HATEROSCEDASTICITY? (x_{p,q}) = 'TRUE' then
                    discard this process to develop a WCEE
                else
                    WCEE_Array [ ] ← ROBUST_REGRESSION (WD)
            else
                if HATEROSCEDASTICITY? (x_{p,q}) = 'TRUE' then
                    WCEE_Array [ ] ← WEIGHT_REGRESSION (WD)
                    WCEE_Array [ ] ← LOGLINEAR_REGRESSION (WD)
                else
                    WCEE_Array [ ] ← MULTIPLE_REGRESSION (WD)
returns WCEE_Array [ ]
```

Figure 44. Algorithm Pseudo-Code for Data Attributes Analysis

Table 17. VIFs on each MOP

MOPs	X1	X2	X3	X7	X8	X9	X10b	X11	X12a	X12b
VIF	3.21	4.59	1.64	1.40	1.64	9.39	9.28	1.45	2.85	4.22

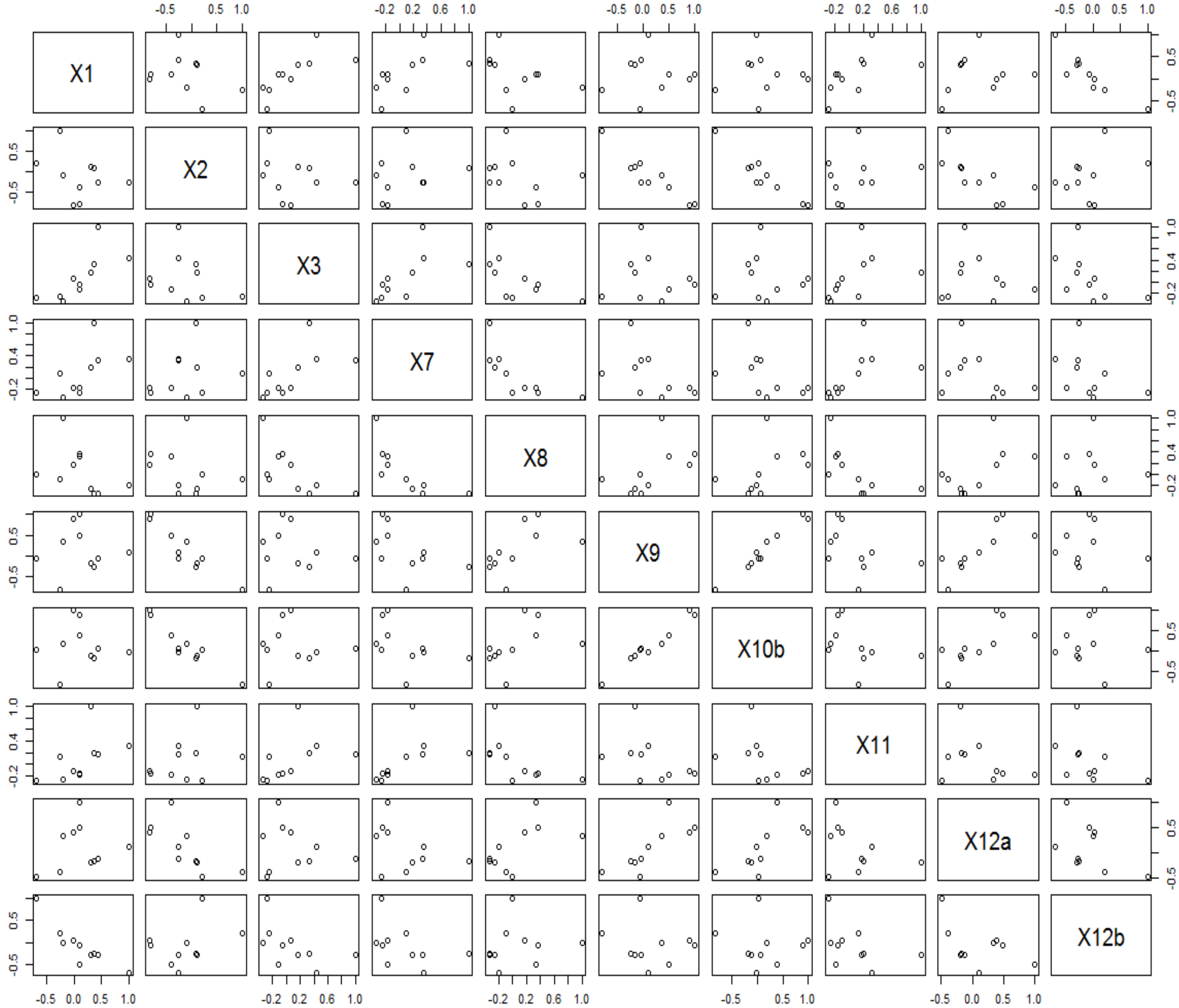

Figure 45. MOPs Scatterplot Matrices

The outliers are identified using Cook's distance which is to estimate the influence of a data point. Figure 46 shows the outliers of the scenarios identified by Cook's distance. The outliers are the nine scenarios which are the 12th, 13th, 39th, 40th, 105th, 129th, 144th, 145th, and 147th scenarios. The existence of the outliers is checked by the relationship between the standardized residuals and the leverage as shown in Figure 47.

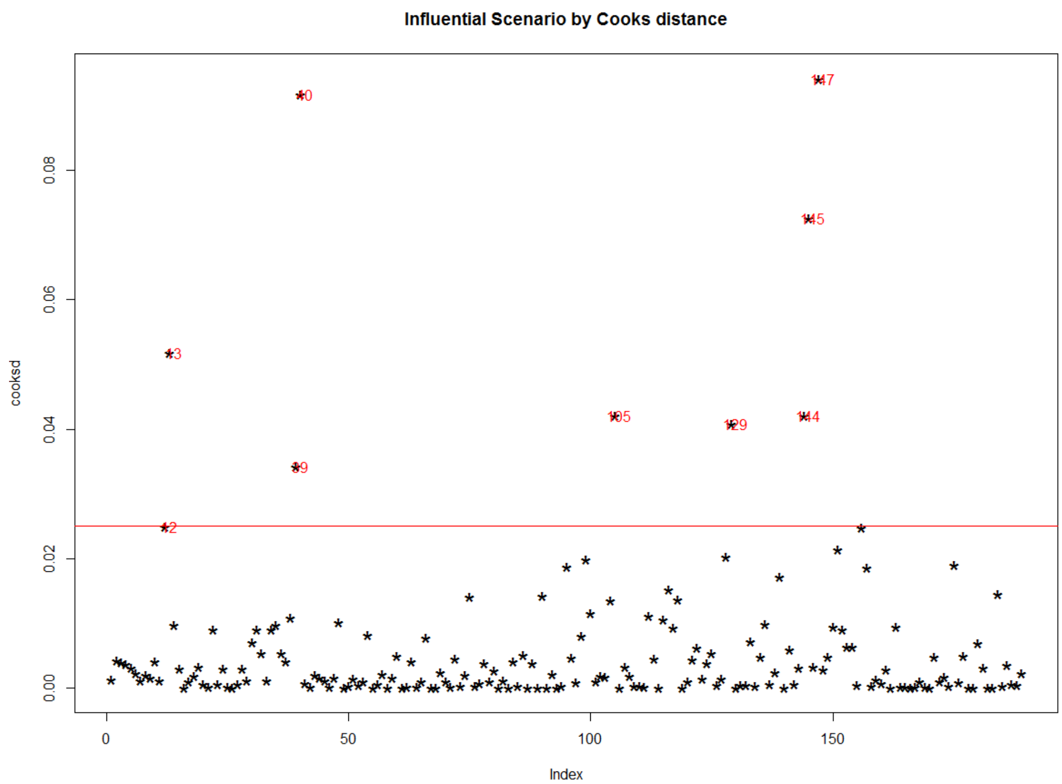

Figure 46. Identification on Outliers of Scenarios by Cook's distance

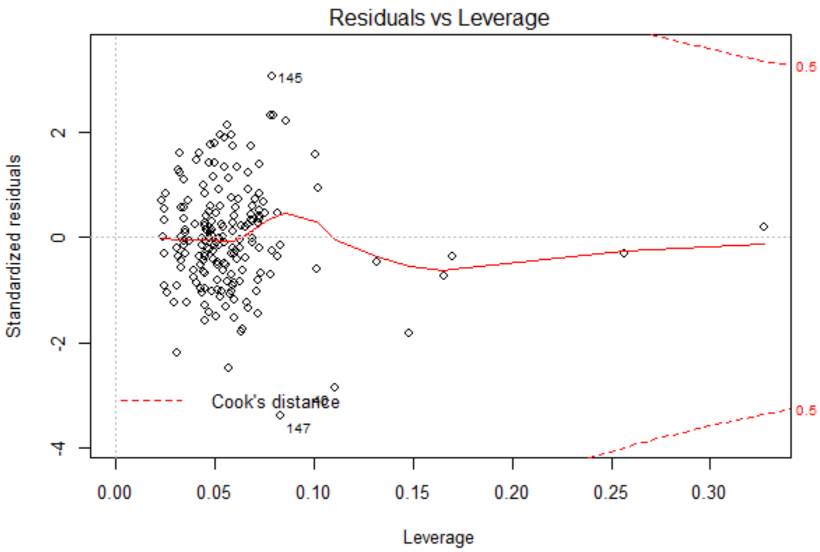

Figure 47. Relationship Plot between Standardized Residuals and Leverage

The heteroscedasticity is analyzed by the relationship between the square root of the standardized residuals and the fitted values. Figure 48 represents the relationship which is not meaningful between the factors. That means there is not heteroscedasticity in a data set.

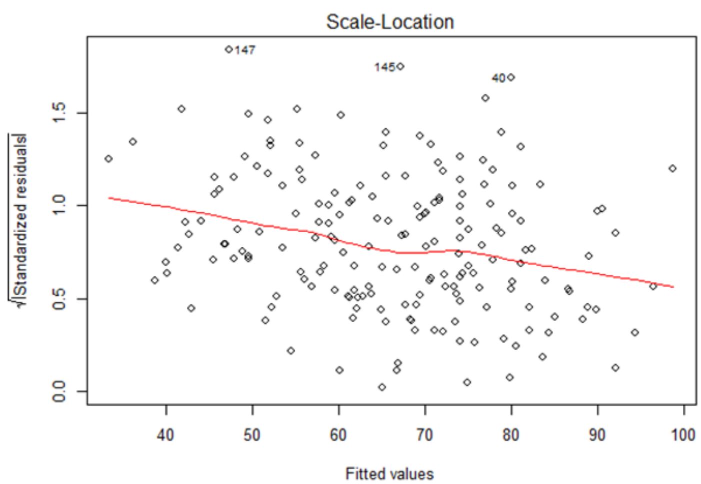

Figure 48. Relationship between Square Root of Standardized Residuals and Fitted Values

To sum up, the outlier exists, whereas the multicollinearity and the heteroscedasticity do not exist as a result of data attributes analysis. The robust regression is used to develop a WCEE for minimizing the outliers' influence. Also, another WCEE is developed using the multiple regression after deleting outliers. Figure 49 shows the data attributes analysis flow.

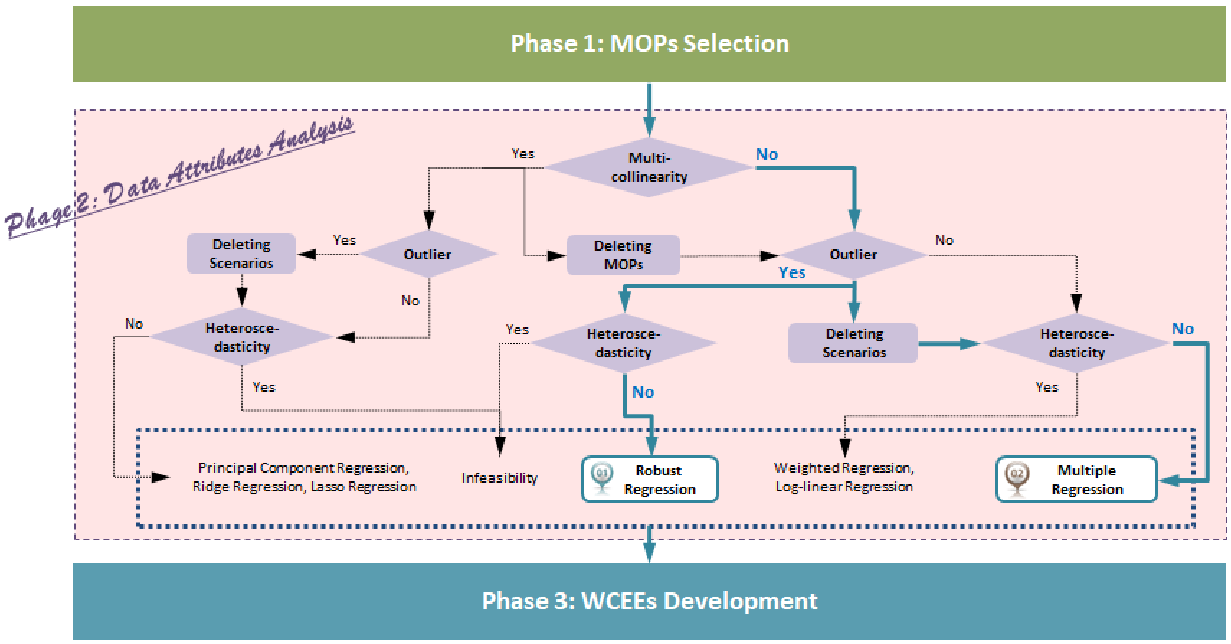

Figure 49. Data Attributes Analysis Flow

4.3.1.1.3 WCEEs development

The WCEEs are built based on the algorithm pseudo-code for the WCEEs development as shown in Figure 50. The first WCEE is developed by the robust regression using an M estimator as shown in Equation (4.18). The statistical evaluations are done on the developed WCEE. Table 18 represents the result of ANOVA. Since the model's p-value in ANOVA is less than 0.001, the first WCEE is statistically significant at an alpha level of 0.05. Also, the estimated coefficients are evaluated by the t-test as shown in Table 19. The entire MOPs' p-values are less than 0.05, and thus, the estimated coefficients are statistically meaningful at an alpha level of 0.05.

```
function WCEEs-DEVELOPMENT (WCEE_Array [ ]) returns Feasible_WCEE_S_Array [ ]
    inputs: WCEE_Array [ ], a data attribute analysis result
        for l = 0 to L do
            for each WCEE_Array [ l : ] do
                If p-value for t-test and F-test of coefficients < 0.05 then
                    Feasible_WCEE_S_Array [ ] ← WCEE_Array [ l : ]
                    If the number of feasible WCEEs ≥ 2 then
                        Feasible_WCEE_S_Array [ ] ← Feasible_WCEE_S_Array [ ] +
                        \ LINEAR_COMBINATION (WCEE_Array [ l : ])
                    else Feasible_WCEE_S_Array [ ]
                else
                    discard WCEE_Array [ l : ]
    returns Feasible_WCEE_S_Array [ ]
```

Figure 50. Algorithm Pseudo-Code for WCEEs Development

$$\hat{Y} = \widehat{39.4151} + \widehat{1.6827}X1 + \widehat{0.2174}X2 - \widehat{0.7562}X3 - \widehat{4.6627}X7 + \widehat{4.2005}X8 - \widehat{0.1530}X9$$
$$+\widehat{19.5601}X10b + \widehat{6.9656}X11 - \widehat{9.3617}X12a - \widehat{8.9166}X12b \qquad (4.18)$$

Table 18. ANOVA on WCEE Developed by Robust Regression

Source	DF	Sum of Sq.	Mean of Sq.	F-Value	p-value
Model	10	31,370.1	3,137.01	24.66	< 0.001
Error	178	22,640.2	127.19		
Total	188	54,010.3			

Table 19. Evaluation of Coefficients Estimated by Robust Regression

MOPs	Value	Std. Error	t-value	p-value
Intercept	39.4151	7.5364	5.23	< 0.001
X1	1.6827	0.675	2.4928	< 0.05
X2	0.2174	0.0463	4.6958	< 0.001
X3	-0.7562	0.3047	-2.4817	< 0.05
X7	-4.6627	0.4967	-9.3883	< 0.001
X8	4.2005	1.5794	2.6596	< 0.01
X9	-0.153	0.0745	-2.0545	< 0.05
X10b	19.5601	5.5483	3.5255	< 0.001
X11	6.9656	0.8181	8.5142	< 0.001
X12a	-9.3617	3.1874	-2.9371	< 0.01
X12b	-8.9166	3.3	-2.702	< 0.01

The second WCEE is built by the multiple regression after deleting nine outliers detected by Cook's distance as shown in Equation (4.19). The developed WCEE is evaluated by ANOVA and the t-test. The result of ANOVA is shown in Table 20. The model's p-value in ANOVA is less than 0.001 and, thus, the second WCEE is statistically meaningful at an alpha level of 0.001. The t-test is done to evaluate the estimated coefficients as shown in Table 21. Since the entire MOPs' p-values are less than 0.05, the estimated coefficients are statistically significant at an alpha level of 0.05.

$$\hat{Y} = \widehat{36.7603} + \widehat{1.7951}X1 + \widehat{0.2154}X2 - \widehat{0.7327}X3 - \widehat{4.8760}X7 + \widehat{3.4721}X8 - \widehat{0.1533}X9 + \widehat{18.0385}X10b + \widehat{7.5378}X11 - \widehat{6.5462}X12a - \widehat{7.4873}X12b \qquad (4.19)$$

Table 20. ANOVA on WCEE Developed by Multiple Regression

Source	DF	Sum of Sq.	Mean of Sq.	F-Value	p-value
Model	10	34,360.1	3,436.01	37.38857	< 0.001
Error	169	15,528.4	91.9		
Total	179	49,888.5			

Table 21. Evaluation of Coefficients Estimated by Multiple Regression

MOPs	Value	Std. Error	t-value	p-value
Intercept	36.76032	6.98141	5.265	< 0.001
X1	1.79505	0.62804	2.858	< 0.01
X2	0.21542	0.04206	5.121	< 0.001
X3	-0.7327	0.27568	-2.658	< 0.01
X7	-4.87595	0.44822	-10.878	< 0.001
X8	3.47209	1.55846	2.228	< 0.05
X9	-0.15333	0.06667	-2.3	< 0.05
X10b	18.03847	4.97297	3.627	< 0.001
X11	7.53784	0.75466	9.988	< 0.001
X12a	-6.54616	2.96561	-2.207	< 0.05
X12b	-7.48729	3.10382	-2.412	< 0.05

The WCEE is developed by the linear combination of the above two WCEEs. The linear combination is done based on the accuracy of the two WCEEs. Each WCEE has weight according to the accuracy index. The weights are given to each WCEE, and then they are added linearly to each other. The PRED(0.3) is used for the accuracy index to develop the linear combination WCEE which is named the ensemble WCEE. The weights are calculated as shown in Equation (4.20). The weights are calculated as shown in Table 22. The ensemble WCEE is shown in Equation (4.21) with the weights.

$$W_k = \frac{\text{PRED}(0.3)_k}{\sum_{k=1}^{2}\text{PRED}(0.3)_k} \tag{4.20}$$

W_k: k_{th} Weight

$$k: \begin{cases} 1: \text{WCEE Developed by the Robust Regression} \\ 2: \text{WCEE Developed by the Multiple Regression} \end{cases}$$

Table 22. Weights for Ensemble WCEE

Type	Robust R. Model	Multiple R. Model
Weight	0.488	0.512

$$\hat{Y} = 36.\widehat{7603} + 1.\widehat{7389}X1 + 0.\widehat{2164}X2 - 0.\widehat{7445}X3 - 4.\widehat{7677}X7 + 3.\widehat{8282}X8 - 0.\widehat{1531}X9$$
$$+18.\widehat{7986}X10b + 7.\widehat{2401}X11 - 7.\widehat{8984}X12a - 8.\widehat{3284}X12b \tag{4.21}$$

4.3.1.1.4 WCEE selection

The best WCEE is selected based on the algorithm pseudo-code for the WCEE selection as shown in Figure 51. The MMRE and PRED(0.3) are calculated to select the best WCEE of the following: a) the WCEE developed by the robust regression, b) the WCEE developed by the multiple regression, and c) the WCEE developed by the linear combination of the above two WCEEs. Figure 52 shows MMRE and PRED(0.3) of three WCEEs. Even though all the calculated MMREs are under 0.25 which is validation criterion, MMRE of the WCEE developed by the multiple regression is the lowest, and thus, the WCEE is selected. Each PRED(0.3) is more than 0.75, and thus, the entire WCEEs are acceptable. However, the WCEEs developed by the multiple regression and the linear combination of two WCEEs have the highest PRED(0.3) and are selected. Therefore, the WCEE developed by the multiple regression is selected as the best WCEE as shown in Table 23.

```
function WCEE-SELECTION (Feasible_WCEE_S_Array [ ]) returns Final_WCEE_List [ ]
    inputs: Feasible_WCEE_S_Array [ ], a data set of weights of feasible WCEEs
        for m = 0 to M do
            for each Feasible_WCEE_S_Array [m: ] do
                If MMRE < 0.25 and PRED(0.30) > 0.75 then
                    Final_WCEE_S_Array [ ] ← Feasible_WCEE_S_Array [m:]
                else
                    discard Feasible_WCEE_S_Array [m:]
                    go to the normalized reduced data step
        Final_WCEE_List [ ] ← The best accuracy model of Final_WCEE_S_Array [ ]
    returns Final_WCEE_List [ ]
```

Figure 51. Algorithm Pseudo-Code for WCEE Selection

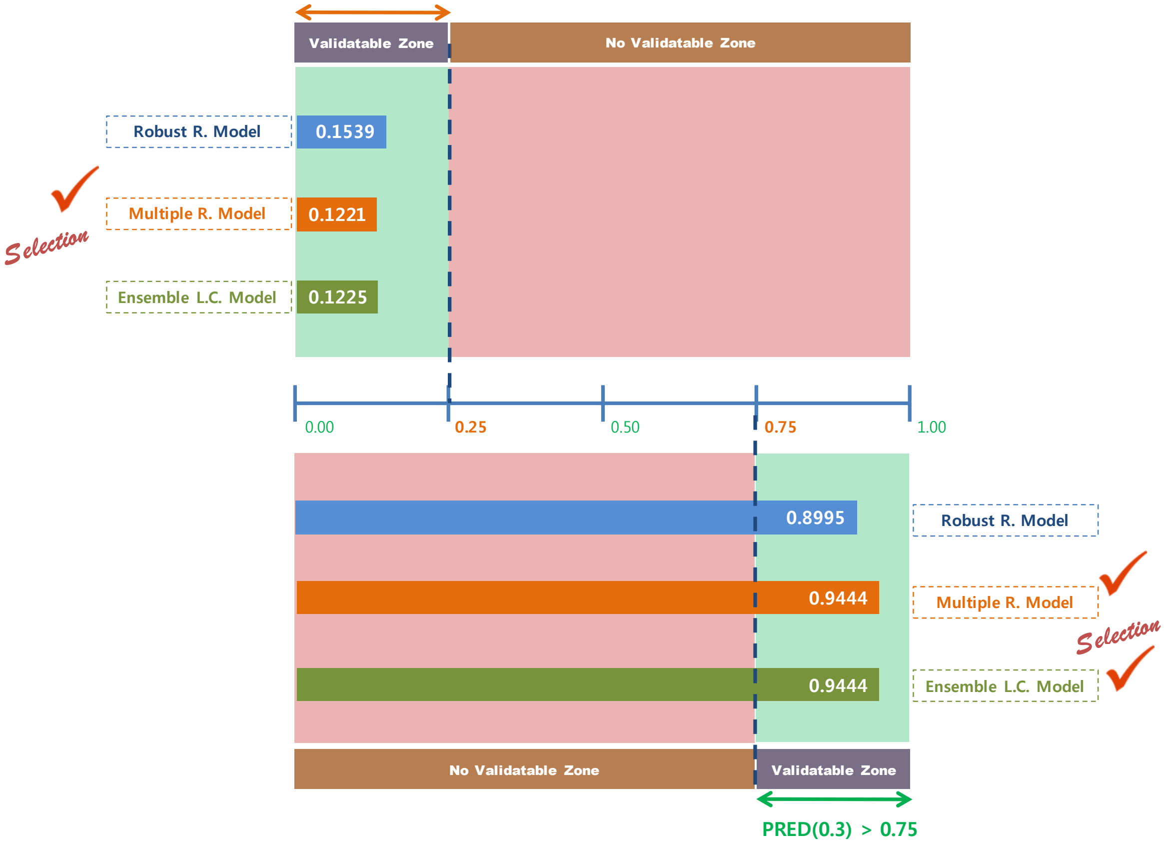

Figure 52. MMRE and PRED(0.3) of three WCEEs

Table 23. Best WCEE Selection

Selection Conditions	Robust R. Model, Equation (4.18)	Multiple R. Model, Equation (4.19)	Ensemble L.C Model, Equation (4.21)
MMRE		×	
PRED (0.3)		×	×
The Final		×	

The MOE and MOPs are finally determined by selecting a final WCEE. Equation (4.22), Equation (4.23), and Equation (4.24) show the estimated MOE's values set, the MOPs' values set, and the estimated coefficients on MOPs. The MOE's values are estimated according to the MOPs' values of scenarios by the developed WCEE.

$$\widehat{\boldsymbol{Y}} = \left\{ \hat{y}_i \middle| \hat{y}_i = \begin{matrix} 44.4951 \\ 42.7 \\ \vdots \\ 77.2763 \end{matrix} \right\} \tag{4.22}$$

$$\boldsymbol{X} = \left\{ x_{i,j} \middle| x_{i,j} = \begin{pmatrix} 1 & 6 & \cdots & 0 \\ 1 & 5 & & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 5 & \cdots & 1 \end{pmatrix} \right\} \tag{4.23}$$

$$\widehat{\boldsymbol{\beta}} = \left\{ \hat{\beta}_j \middle| \hat{\beta}_j = \begin{matrix} 36.7603 \\ 1.7951 \\ \vdots \\ -7.4873 \end{matrix} \right\} \tag{4.24}$$

$\widehat{\boldsymbol{Y}}$: Estimated MOE Values Set

$\boldsymbol{X}$: MOPs' Values Set

$\widehat{\boldsymbol{\beta}}$: Estimated Coefficients Set

$\hat{y}_i$: Estimated Value on i_{th} Simulation

$x_{i,j}$: Value on i_{th} Simulation and j_{th} MOP, (i = 1, …, 179, j = 0, 1, 2, 3, 7, 8, 9, 10b, 11, 12a, 12b)

$x_{i,j=0}$: Value for MOE's Intercept Calculation on i_{th} Simulation

$\hat{\beta}_j$: Estimated Coefficient on j_{th} MOP (j = 0, 1, 2, 3, 7, 8, 9, 10b, 11, 12a, 12b)

$\hat{\beta}_{j=0}$: MOE's Intercept

4.3.1.2 MOPs' optimal values

The MOPs' optimal values are considered as the optimization problem. MIP is modeled to find the best MOP's values for the developed WCEE and constraints. The objective equation is shown in Equation (4.25). The constraints are represented in Equation (4.26), Equation (4.27), Equation (4.28), Equation (4.29), Equation (4.30), Equation (4.31), Equation (4.32), and Equation (4.33). The constraints are extracted in Equation (4.9), Equation (4.10), and Equation (4.11) by the expert system.

$$\begin{aligned} \text{Maximize} \quad & f(x_1, \ldots, x_k) \\ = \; & 36.7603 + 1.7951x_1 + 0.2154x_2 - 0.7327\,x_3 - 4.8760x_7 + 3.4721x_8 - 0.1533x_9 \\ & + 18.0385x_{10b} + 7.5378x_{11} - 6.5462x_{12a} - 7.4873x_{12b} \end{aligned} \tag{4.25}$$

Subject to

$$1 \leq x_1 \leq 5 \tag{4.26}$$

$$6 \leq x_2 \leq 10 \tag{4.27}$$

$$5 \leq x_3 \leq 20 \tag{4.28}$$

$$3 \leq x_7 \leq 5 \tag{4.29}$$

$$1 \leq x_8 \leq 10 \tag{4.30}$$

$$10 \leq x_9 \leq 50 \tag{4.31}$$

$$1 \leq x_{11} \leq 3 \tag{4.32}$$

$$0.17x_1 + 18.8x_{11} \leq 52 \tag{4.33}$$

$x_{10b}, x_{12a}, x_{12b} \in \{0, 1\}$

$\forall x_1, x_2, x_7, x_{11} \in \mathbb{Z}$

$\forall x_3, x_8, x_9 \in \mathbb{R}$

k = 1, 2, 3, 7, 8, 9, 10b, 11, 12a, 12b

The MOPs' optimal values are searched based on the above modeled MIP using POM-QM software (http://wps.prenhall.com/bp_weiss_software_1/1/358/91660.cw/index.html). The MOE and the MOPs' optimal values are found as shown in Equation (4.34). The MOE can be achieved up to 95.90% survival rate under the given MOPs.

$$\boldsymbol{V}^* = \begin{bmatrix} y^* \\ {x_1}^* \\ {x_2}^* \\ {x_3}^* \\ {x_7}^* \\ {x_8}^* \\ {x_9}^* \\ {x_{10b}}^* \\ {x_{11}}^* \\ {x_{12a}}^* \\ {x_{12b}}^* \end{bmatrix} = \begin{bmatrix} 95.90 \\ 5 \\ 10 \\ 5.00 \\ 3 \\ 10.00 \\ 10.00 \\ 1 \\ 2 \\ 0 \\ 0 \end{bmatrix} \qquad (4.34)$$

$\boldsymbol{V}^*$: A Data Set of the Optimal MOE and MOPs

y: The Optimal MOE

${x_i}^*$ = The Optimal MOP, (*i* = 1, 2, 3, 7, 8, 9, 10b, 11, 12a, 12b)

4.3.1.3 Verification and validation for the data

The SMEs' reviews are used to validate the normalized data transformed from the reduced data by the normalization process. The normalized data is checked using five-point Likert Scale by SMEs. The assessment-level of four and half points is larger than the requirement-level of three points, and thus, the normalization data is accepted.

The WCEE is built using the WCEE development algorithm by R programming language. The WCEE development phase is verified by debugging the code on whether the algorithm is implemented correctly to R programming language. It is accepted based on the verification results. Also, the optimal values' estimation phase is verified using the traceability assessment method when the objective function and constraints are implemented to the POM-QM software.

The developed WCEE is validated using the hypothesis testing and prediction testing. The t-test and ANOVA are used for the hypothesis and the MMRE and PRED(0.3) are applied for prediction testing. The optimal values are validated by the data analysis. The sensitivity analysis is used for the data analysis. The developed WCEE and optimal values are validated based on the testing results.

4.3.2 STEP 5: Reporting Results using Visualization

The values for the probability map of the MOE are calculated as shown in Equation (4.35). The calculated results' matrix is shown as shown in Figure 53. The blue axis shows the blue team's MOPs, whereas the red axis represents the red team's MOPs. The matrix format enables decision makers to compare their team's each MOP with all the opposite team's MOPs. They can find the most influential MOP against the opposite team's MOPs without complex equations. When the x_{11} is 3, all values of the x_1 do not satisfy the constraint which is shown in Equation (4.26) and Equation (4.33). Therefore, it is an infeasible area which is represented as a black part in Figure 53.

$$\mathrm{MOE}_{x_i,x_j} = f\left(x_i, x_j, {}^{\forall}\, {x_k}^* \,\middle|\, x_{k\neq i,j} = {x_k}^*\right) \tag{4.35}$$

MOE_{x_i,x_j}: MOE Value by x_i, x_j, and $^{\forall}\ {x_k}^*$

$x_i,\ x_j,\ x_k$: Value of $x_i,\ x_j$, and x_k

${x_k}^*$: The Optimal Value of x_k to Maximize the MOE

i, j, k = 1, 2, 3, 7, 8, 9, 10b, 11, 12a, 12b

		X3				X7			X9					X12a		X12b	
		5	10	15	20	3	4	5	10	20	30	40	50	0	1	0	1
X1	1	88.72	85.06	81.39	77.73	88.72	83.84	78.97	88.72	87.19	85.65	84.12	82.59	88.72	82.17	88.72	81.23
	2	90.52	86.85	83.19	79.52	90.52	85.64	80.76	90.52	88.98	87.45	85.92	84.38	90.52	83.97	90.52	83.03
	3	92.31	88.65	84.98	81.32	92.31	87.43	82.56	92.31	90.78	89.24	87.71	86.18	92.31	85.76	92.31	84.82
	4	94.11	90.44	86.78	83.11	94.11	89.23	84.35	94.11	92.57	91.04	89.51	87.97	94.11	87.56	94.11	86.62
	5	95.90	92.24	88.57	84.91	95.90	91.02	86.15	95.90	94.37	92.83	91.30	89.77	95.90	89.35	95.90	88.41
X2	6	95.04	91.38	87.71	84.05	95.04	90.16	85.29	95.04	93.51	91.97	90.44	88.91	95.04	88.49	95.04	87.55
	7	95.25	91.59	87.93	84.26	95.25	90.38	85.50	95.25	93.72	92.19	90.66	89.12	95.25	88.71	95.25	87.77
	8	95.47	91.81	88.14	84.48	95.47	90.59	85.72	95.47	93.94	92.40	90.87	89.34	95.47	88.92	95.47	87.98
	9	95.69	92.02	88.36	84.69	95.69	90.81	85.93	95.69	94.15	92.62	91.09	89.55	95.69	89.14	95.69	88.20
	10	95.90	92.24	88.57	84.91	95.90	91.02	86.15	95.90	94.37	92.83	91.30	89.77	95.90	89.35	95.90	88.41
X8	1	64.65	60.99	57.32	53.66	64.65	59.78	54.90	64.65	63.12	61.59	60.05	58.52	64.65	58.11	64.65	57.16
	2	68.12	64.46	60.80	57.13	68.12	63.25	58.37	68.12	66.59	65.06	63.52	61.99	68.12	61.58	68.12	60.64
	3	71.60	67.93	64.27	60.61	71.60	66.72	61.84	71.60	70.06	68.53	67.00	65.46	71.60	65.05	71.60	64.11
	4	75.07	71.40	67.74	64.08	75.07	70.19	65.32	75.07	73.53	72.00	70.47	68.94	75.07	68.52	75.07	67.58
	5	78.54	74.88	71.21	67.55	78.54	73.66	68.79	78.54	77.01	75.47	73.94	72.41	78.54	71.99	78.54	71.05
	6	82.01	78.35	74.69	71.02	82.01	77.14	72.26	82.01	80.48	78.95	77.41	75.88	82.01	75.47	82.01	74.52
	7	85.48	81.82	78.16	74.49	85.48	80.61	75.73	85.48	83.95	82.42	80.89	79.35	85.48	78.94	85.48	78.00
	8	88.96	85.29	81.63	77.97	88.96	84.08	79.20	88.96	87.42	85.89	84.36	82.82	88.96	82.41	88.96	81.47
	9	92.43	88.76	85.10	81.44	92.43	87.55	82.68	92.43	90.90	89.36	87.83	86.30	92.43	85.88	92.43	84.94
	10	95.90	92.24	88.57	84.91	95.90	91.02	86.15	95.90	94.37	92.83	91.30	89.77	95.90	89.35	95.90	88.41
X10b	0	77.86	74.20	70.53	66.87	77.86	72.99	68.11	77.86	76.33	74.80	73.26	71.73	77.86	71.32	77.86	70.37
	1	95.90	92.24	88.57	84.91	95.90	91.02	86.15	95.90	94.37	92.83	91.30	89.77	95.90	89.35	95.90	88.41
X11	1	88.36	84.70	81.04	77.37	88.36	83.49	78.61	88.36	86.83	85.30	83.76	82.23	88.36	81.82	88.36	80.88
	2	95.90	92.24	88.57	84.91	95.90	91.02	86.15	95.90	94.37	92.83	91.30	89.77	95.90	89.35	95.90	88.41
	3																

Figure 53. MOE values based on MOPs

Figure 54 shows the probability map of the MOE, which is designed from values of Figure 53, for decision makers to make decisions effectively and efficiently. Decision makers can visually identify that the x_8, x_{10b}, and x_{11} are more important MOPs than x_1, and x_2, because a MOP with the sudden color change means more influential than the others.

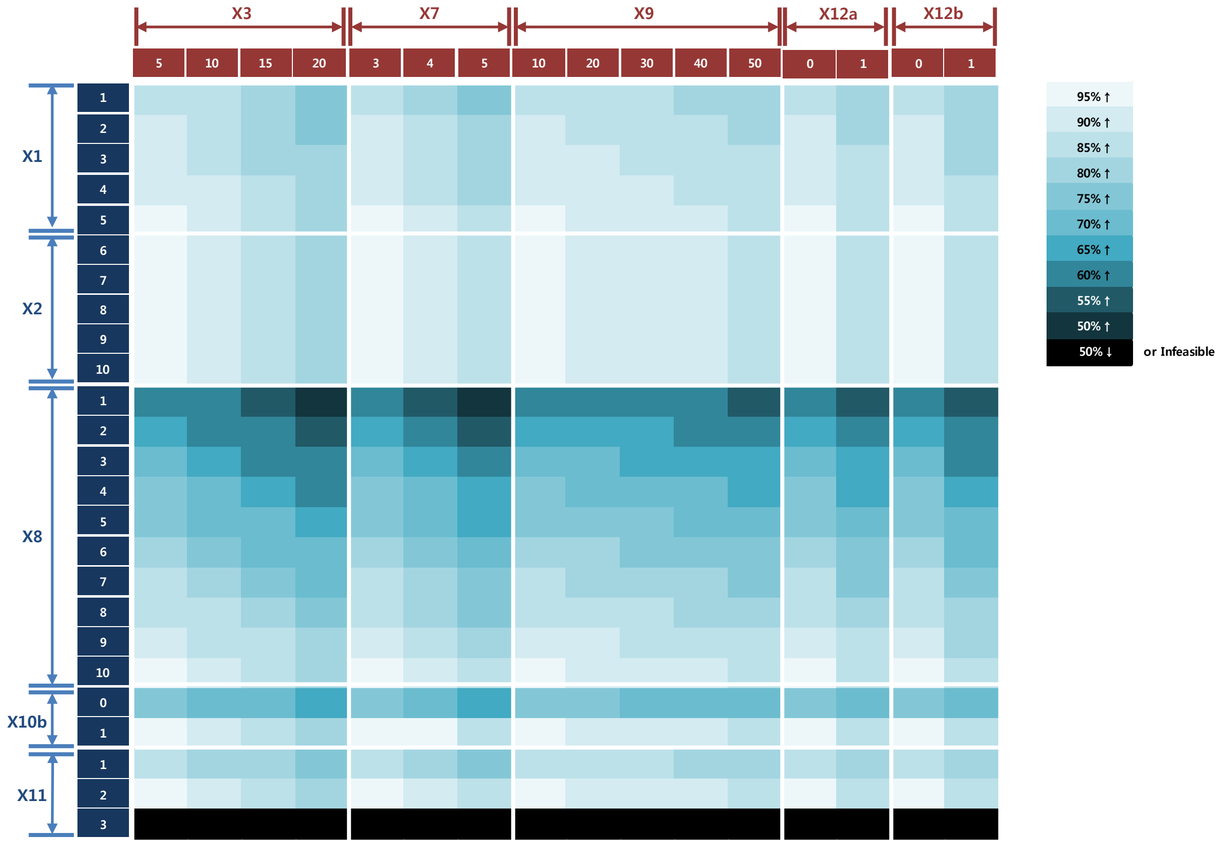

Figure 54. Probability Map

The MOE index is developed by Equation (4.25) under constraints. The MOE index matrix shows the priority of combination of MOPs to maximize the MOE as shown in Figure 55. The values of MOPs, which have values of a real number, are considered as the integer type. This is because it is not possible to calculate all values of a real number.

Effectiveness Index	Y	X1	X2	X3	X7	X8	X9	X10b	X11	X12a	X12b
1	95.90	5	10	5	3	10	10	1	2	0	0
2	95.75	5	10	5	3	10	11	1	2	0	0
3	95.69	5	9	5	3	10	10	1	2	0	0
4	95.59	5	10	5	3	10	12	1	2	0	0
5	95.53	5	9	5	3	10	11	1	2	0	0
6	95.47	5	8	5	3	10	10	1	2	0	0
7	95.44	5	10	5	3	10	13	1	2	0	0
8	95.38	5	9	5	3	10	12	1	2	0	0
9	95.32	5	8	5	3	10	11	1	2	0	0
10	95.29	5	10	5	3	10	14	1	2	0	0
⋮	⋮					⋮					
1457	90.00	4	10	6	3	10	32	1	2	0	0
1458	90.00	5	6	10	3	10	19	1	2	0	0
⋮	⋮					⋮					
51169	80.00	5	9	8	4	10	17	1	1	0	0
51170	80.00	5	7	7	3	7	32	1	2	0	0
⋮	⋮					⋮					
155797	75.00	3	6	5	4	9	14	1	2	0	1
155798	75.00	4	10	9	4	10	41	1	2	1	0
⋮	⋮					⋮					
366721	70.00	2	7	9	3	9	49	1	2	0	1
366722	70.00	5	10	11	4	8	24	1	1	0	0
⋮	⋮					⋮					
7872000	0.00	1	6	20	5	1	50	0	1	1	1

Blue Team MOPs Red Team MOPs 95↑ 90 85 80 75 70 65 60 55 50 infeasible

Figure 55. Effectiveness Index on the MOE and MOPs

The final document for reporting results is composed of the MOE, MOPs, WCEE, a probability map matrix, a MOE index matrix, etc. It helps to drive decision makers to comprehensive understandings. If decision makers judge that the final document has defects, insufficient contents, uncertain results, latest information, etc., they can order for the analysts to reanalyze the WCE based on the final document. This part is omitted in this case study.

The F-16 MOE estimator is developed and provided for decision makers to easily get more detailed results as shown in Figure 56. It helps the MOE to be recalculated by the estimator according to the changed constraints. That means it is a very flexible tool. The estimator can also play a role as an additional scientific staff when decision makers make an important decision.

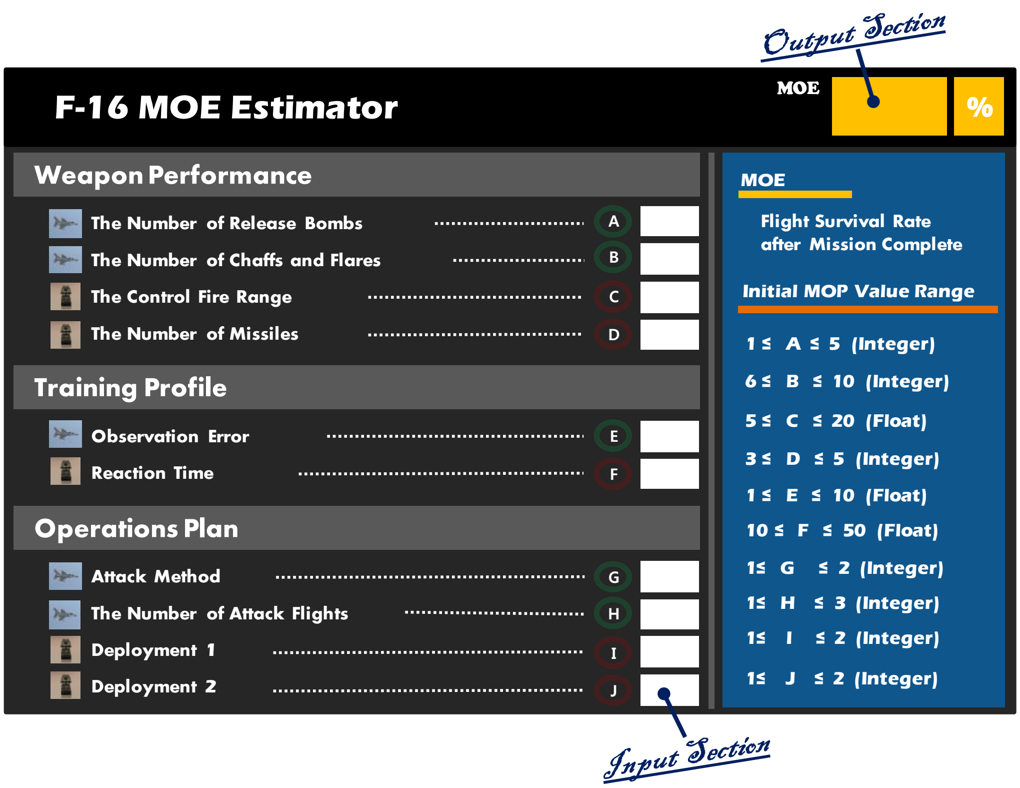

Figure 56. F-16 MOE Estimator Format

CHAPTER FIVE: EXTENDABLE AREAS

This chapter shows extendable areas using the suggested methodology on the WCE analytics. Each area is explained with examples based on the probability map developed by the estimated WCEE. This chapter represents the potential power on the new methodology on the WCE analytics.

5.1 Overview

The suggested methodology for the WCE analytics can be utilized to various areas. Figure 57 represents the extendable areas which are the following: a) supporting operations analytics and plan, b) guiding effectiveness-focused training, and c) recommending ROC for weapon acquisition. The suggested methodology can be flexibly applied by focusing on their areas' purpose. This remaining section of this chapter shows an example for three extendible areas just based on the developed WCEE.

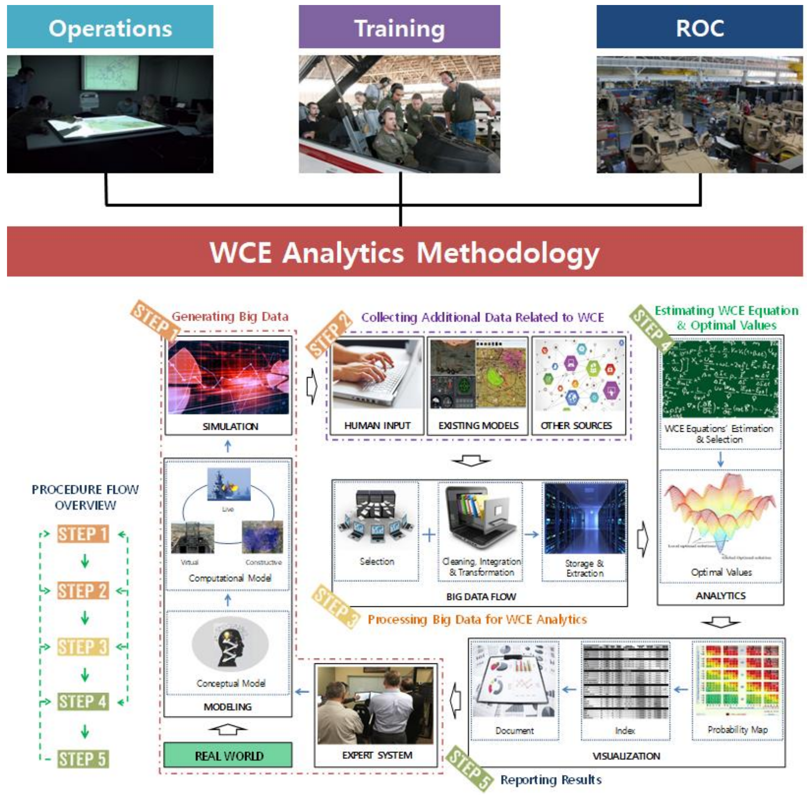

Figure 57. Extendable Areas Using Suggested WCE Analytics Methodology

5.2 Supporting Operations Analytics and Plan

Although the number of staff can vary for every a type of military unit, several staff does exist to support the command in each military unit. They have a variety of tasks, but important one is to give the command information for the operations analytics and plan. The command makes a decision based on the information offered by the staff. Since the information is given by the staff that has human factor errors, the information can be subjective and inaccurate. At this time, the information given by WCE analytics can play an important role. The information can be more objective and accurate than the one given by the staff. This is because the WCEE is statistically developed through the WCE analytics based on a variety of scenarios. Therefore, the WCEE can be a scientific staff to support the command decisions in the military unit.

The operations plan is affected by the 6 MOPs of 10 MOPs, which are the number of released bombs, F-16 flights, and SA-8 missiles as well as the F-16 attack and SAM deployment method. Therefore, the 6 MOPs are selected from the WCEE as shown in Figure 58. The operations analytics is necessary to give information for the operations plan. The MOE probability map, which is developed based on the WCEE, is used as an important tool for the operations analytics. Figure 59 shows the MOE probability map based on the MOPs influencing the operations plan. It enables the command to decide the best operations alternative effectively and efficiently under given conditions. The MOP probability map can be analyzed in various ways according to conditions. For example, without the information on the SAM, the type of attack methods is the most important MOP. This is because the MOE is the most sharply changed according to the MOP on the attack method. There are three attack methods in Figure 60. The attack method "B", which is X_{10b} is 1, is the best option to maximize the MOE. Therefore, the attack method "B" can be considered as the first priority in the operations plan by the command.

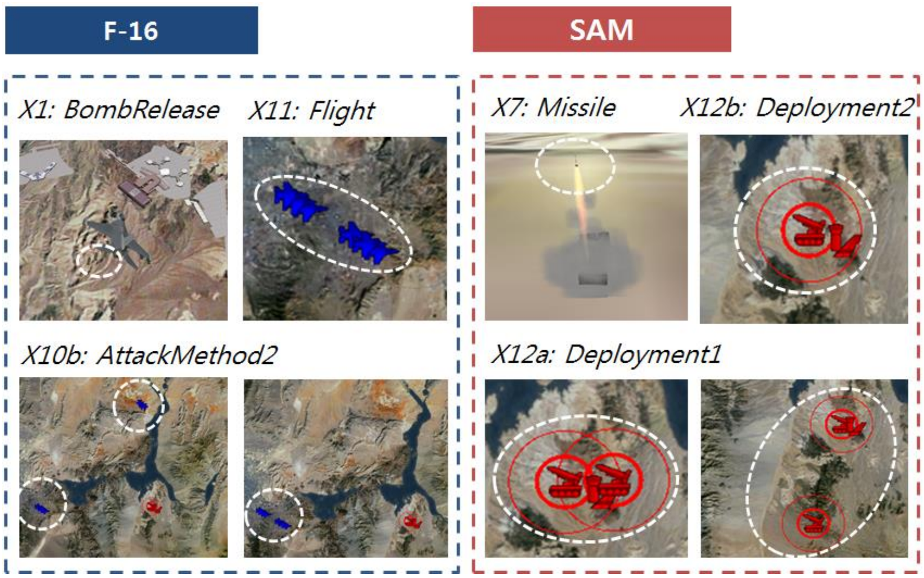

Figure 58. MOPs for Operations Analytics and Plan

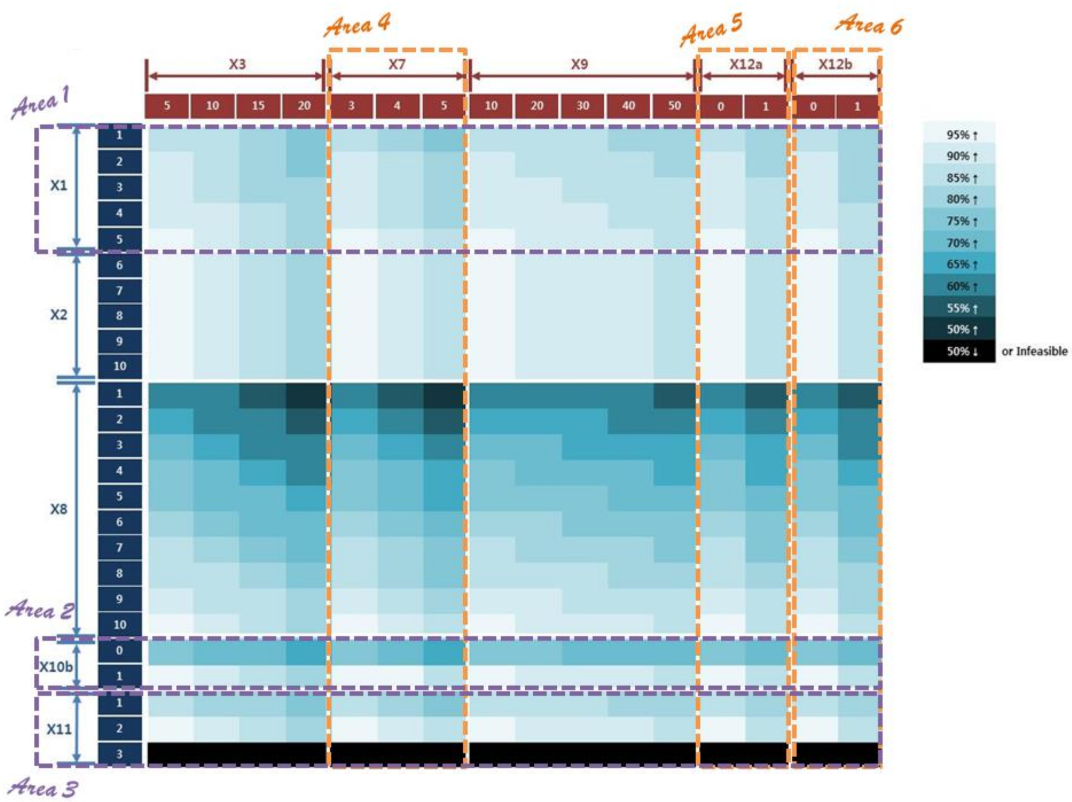

Figure 59. MOE Probability Map based on MOPs Influencing Operations Plan

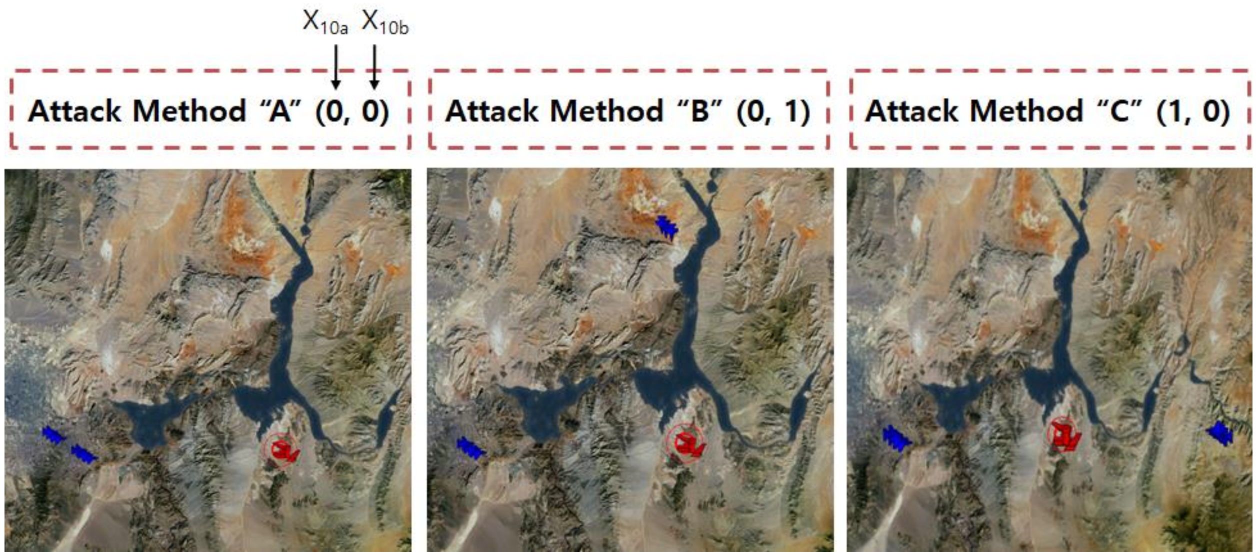

Figure 60. Three Attack Methods of F-16s

5.3 Guiding Effectiveness-Focused Training

The human factors are important in a weapon system. This is because the WCE can be various results according to the ability of the weapon operator. Therefore, training for the weapon operator is a significant factor to maximize the WCE. However, most training levels are decided by empirical factors instead of the scientific ones. The scientific method is necessary to train the weapon operator. This is because the training level decided by the scientific method can give the effectiveness-focused training standards for the weapon operator to maximize the WCE. The weapon operator can be trained effectively and efficiently when based on the training standards. The method is the suggested new methodology for the WCE analytics. The methodology can help to decide the scientific levels for the effectiveness-focused training of the weapon operator.

The two MOPs, which are the observation error and the reaction time, are related to human factors for the weapon operator training as shown in Figure 61. The two MOPs are analyzed, and then, the objective training levels of the weapon operator are decided to maximize the WCE based on the analyzed results. The MOE probability map is shown based on the two MOPs in Figure 62. The objective training levels of the weapon operator can be decided on the observation error and the reaction time. However, the objective training levels can be different according to given conditions. The conditions can be the human ability, cost effectiveness, mission goal, etc. For example, when the observation error is enhanced by a unit, the MOE is gradually increased. However, the weapon operator is not needed to be trained until the observation error of 10 level is reached, if the mission's goal is to keep the MOE values of more than 75. If the observation error is achieved to 8 level, the mission's goal can be completed. Therefore, the objective training level for the observation error is decided to 8 level. The standard enables the decision maker to reduce the unnecessary efforts.

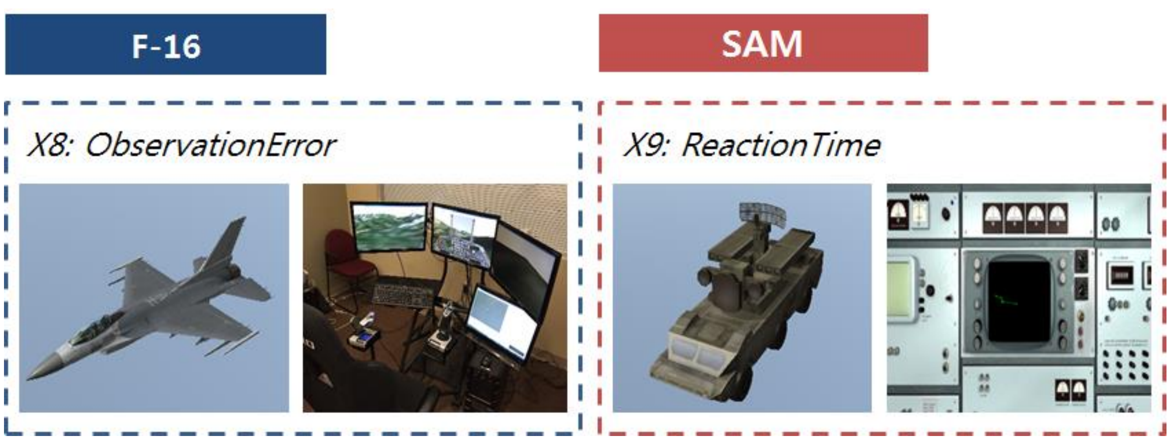

Figure 61. MOPs related to Human Factors for Weapon Operator Training

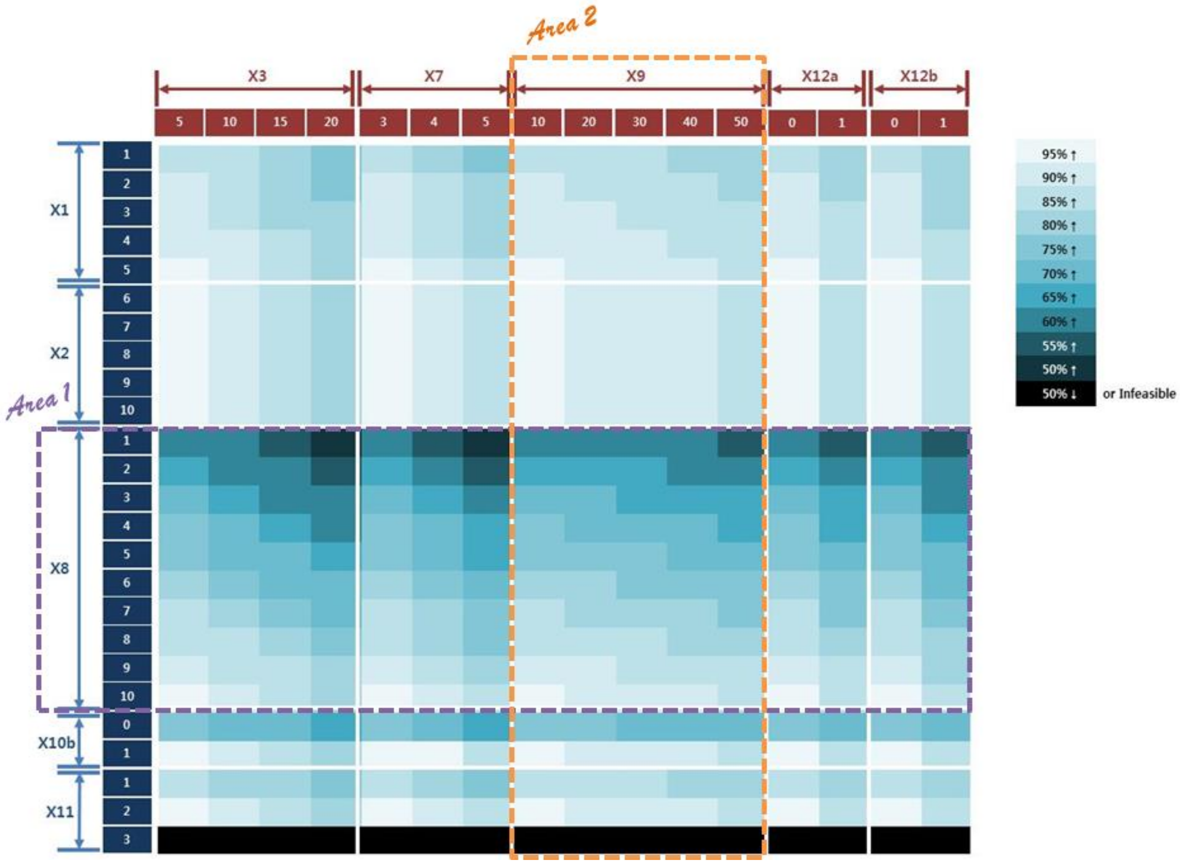

Figure 62. MOE Probability Map based on MOPs related to Human Factors for Weapon Operator Training

5.4 Recommending ROC for Weapon Acquisition

The Required Operational Capability (ROC) is significant for the weapon acquisition. This is because the weapon is developed by the guidelines supplied by the ROC. If the ROC is not properly decided, the weapon developed by the ROC cannot be effective. The developed weapon can be useless under various operational scenarios due to the underestimated capability, or the weapon can be developed and operated more than necessary due to the overestimated capability. The first case can result in mission failure and the second case can lead to excessive cost during the weapon life cycle. Therefore, the ROC is needed to be decided through experiments under various scenarios. However, the decided ROC does not consider various operational situations until recently. Therefore, the suggested methodology can play a role to solve the problem.

The ROC is affected by the 3 MOPs of 10 MOPs, which are the number of released bombs, chaffs and flares, and control fire range as shown in Figure 63. The MOE probability map created by the WCEE is used as an important tool for the ROC recommendation. Figure 64 shows the MOE probability map used for recommending the ROC. For example, the number of chaffs and flares is not the important MOP for the ROC recommendation. This is because the color representing the MOE level is not changed according to the number of chaffs and flares, although the MOE actually is a little bit changed. However, the number of bomb release has more influence for the MOE, because the MOE level is changed according to the number of released bombs. Therefore, the MOP is needed to be decided by given conditions. If, as a condition, the mission's goal is to keep the MOE values of more than 80, the ROC of 3 is recommended.

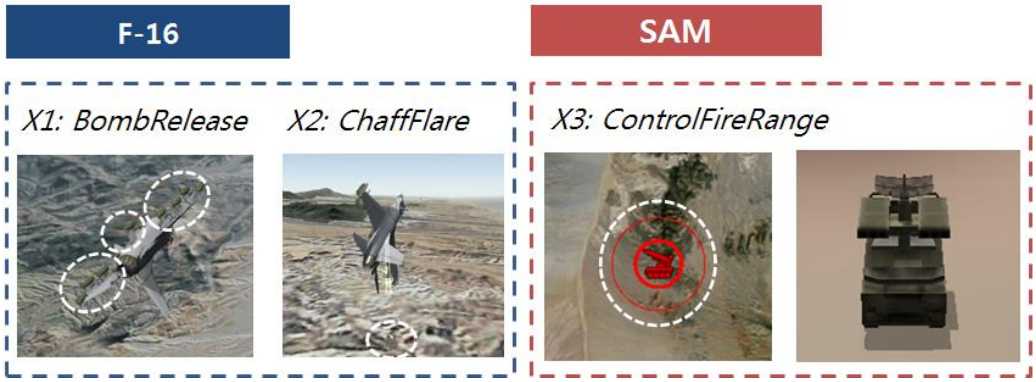

Figure 63. MOPs for Recommending ROC

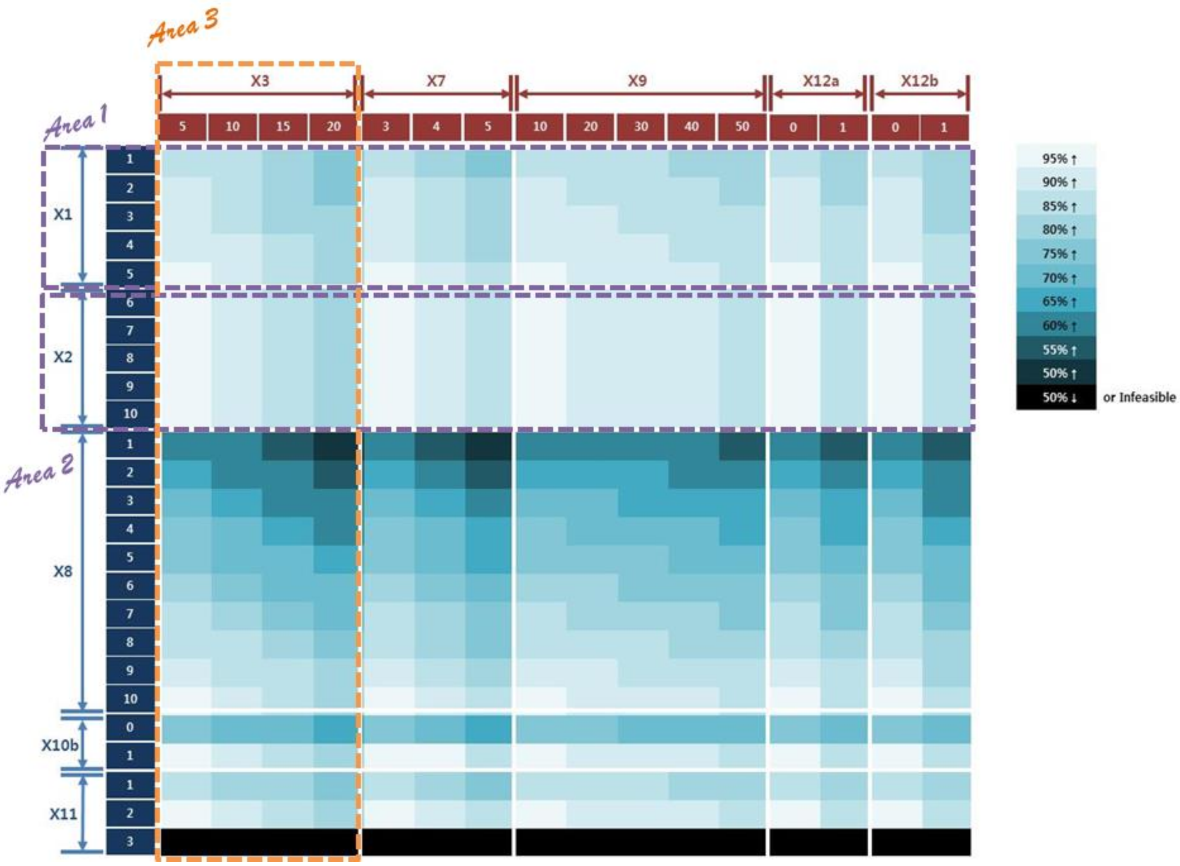

Figure 64. MOE Probability Map based on MOPs related to ROC Recommendation

CHAPTER SIX: CONCLUSIONS

This chapter concludes this research. The research is briefly summarized, and the contributions obtained from this research are suggested. Also, the limitations, which this research has, are identified and the future works are suggested to overcome these limitations.

6.1 Summary

Chapter 1 introduces the background of this research. It includes the importance of the WCE analytics and the limitations represented when the WCE is analyzed in the real world. Also, this chapter identifies the problems that the existing WCE analytics methodology using DM&S has. An approach, which uses big data generated by LVC simulations, is suggested to solve the problems. The purpose of this research is suggested, and then, the goal is developed to satisfy the purpose. The nine objectives are constructed to achieve the goal. The three potential contributions are stated.

Chapter 2 classifies the existing research related to the WCE analytics into three viewpoints, which are the reliability, efficiency, and economics. The computer based DM&S, which is related to the generation of big data, is explained. The LVC simulations are represented and differentiated based on the types of personnel, systems, and operation. The principal levels of the military models, which are the engineering, engagement, mission, and theatre levels, are explained. The various SSAs are identified and explained based on the historical relationship, the comparison of characters, and the usage percentage. The V&V definition and process is represented for the existing research. The V&V methods are classified based on informal testing and formal testing, and static testing and dynamic testing. The general big data and defense big data are explained separately. The big data definition, characteristics, analysis techniques, and technology are shown in the general big data. Examples related to defense big data are explained based on the current existing research.

Chapter 3 suggests the new methodology for the WCE analytics using big data generated by the LVC simulations. The procedure on the suggested methodology is introduced and the WCE analytics system diagram and formalism are explained to implementation of the procedure. The methodology is suggested according to five steps which are a) generating big data, b) collecting additional data related to the WCE, c) processing big data for WCE analytics, d) estimating WCE equation and optimal values, and e) reporting results. The first step, generating big data, explains the development processes for the conceptual and computational models for generating big data. The second step, collecting additional data related to the WCE, represents how to collect data, what data to collect, where to collect data, and who to collect data. The third step, processing big data for WCE analytics, shows three processes, which are a) selection, b) cleaning, integration, and transformation, and c) storage and extraction to prepare data to estimate the WCEE. The fourth step, estimating the WCE equation and optimal values, explains an algorithm for the WCEE development using a flowchart. Also, the optimization process using MIP is explained. The fifth step, reporting results, shows the probability map and effectiveness index examples to visualize the WCE analytics results. The factors for the document, which is the reporting result, are suggested.

Chapter 4 shows the case study on the F-16s CE analytics against the SAMs using the ULM for applying the suggested methodology. The pilot model is developed, implemented, and analyzed for the case study. This case study is done in two sections, which are a) data preparation for the F-16 CE analytics by LLM_{DP}, and b) F-16 CE analytics by LLM_{AN} and the visualization. The big data is generated by the VC simulations using the developed computational model, which consists of the SIMbox and the VR-Forces. The constraints data is collected as the additional data. The generated and collected big data are processed to extract the reliable and necessary data for the F-16 CE analytics by using the HDFS and the MapReduce process on a JVM. The F-16 CE equation is developed and evaluated through the WCEE development algorithm using R programing language. The optimal values influencing the F-16 CE are found using the POM-QM software for MIP. The F-16 analytics results are reported using visualization methods which are the probability map and effectiveness index. Also, the MOE estimator is developed for use as a scientific staff under various conditions.

Chapter 5 shows the extendable areas using the suggested methodology. The extendable areas are explained by suggesting examples based on the developed WCEE. The extendable areas are a) supporting operations analytics and plan, b) guiding effectiveness-focused training, and c) recommending ROC for weapon acquisition. The probability map, which is developed by the estimated WCEE, is used to show examples in each extendable area.

6.2 Contribution

This dissertation has three contributions which are the following: (a) initially composing the basic knowledge for big data and DM&S related to the WCE analytic which can serves as the stepping-stone for future research, (b) proposing the new methodology for the WCE analytics which can solve the problems of the existing research related to the WCE, and (c) suggesting the extendable areas which can apply to the suggested new methodology.

This dissertation overviews WCE, DM&S, and DBD, establishes the WCE classification, identifies the research challenges, and investigates and suggests solutions to overcome the research challenges from the viewpoint of the first contribution. Furthermore, the suggested new methodology as a new concept results in higher fidelity for the WCE analytics than the existing methodology from the viewpoint of the second contribution. This is because it considers and analyzes a variety of MOPs by using big data techniques and technology, and LVC simulations. The new methodology shows benefits as the following: (a) utility of abundant data, a-1) LVC simulations benefits utility, a-2) big data technique benefits utility, and a-3) external source utility, (b) modeling reality; b-1) assumption degree minimization, b-2) various factors application, and b-3) various scenario applications, (c) generalizations; c-1) application flexibility, and c-2) comprehensive analysis, (d) analytics result usability; d-1) WCE equation estimation with various factors, and d-2) optimal values recommendation based on constraints. Lastly, from the viewpoint of the third contribution, the new methodology can be applied to various areas as following: (a) supporting operations analytics and plan, (b) guiding effectiveness-focused training, and (c) recommending ROC for the weapon acquisition.

6.3 Limitations and Future Work

There are four limitations identified from this research. First, the WCE analytics using the new methodology is a costly and time-consuming process compared to the existing methodology. This is because the new methodology uses live and virtual simulation models including human factors instead of only using the constructive simulation model. The more VC simulation models are connected to enhance the fidelity, the more required the cost and time are for the WCE analytics. Second, the WCEE development algorithm has a limitation for the development of the best WCEE for all cases. The generated and collected data has various characteristics according to each case, so the specific algorithm in all cases cannot be the best for the WCEE development. That means that the absolute algorithm cannot exist. Third, the new methodology requires much time to manually show various scenarios. Each scenario should be implemented to the computational model for generating big data through simulations. This step is time-consuming work. Fourth, this dissertation does not fully explain the extendable areas using the new methodology. This research focuses on the WCE analytics, and thus, it has problems in covering all the extendable areas.

Four future works are suggested to solve the limitations. First, a general model should be developed for offering all the training and analytics functions. If the training model and the analytics model are separately developed, it is an inefficient approach. Participation of many operators is not effective and efficient just for the WCE analytics. If the big data generated by the training model is not utilized for the WCE analytics, also it is not effective and efficient. Therefore, the general model, which is equipped with the training and analytics functions, should be developed, and validated and verified according to the model's intended use. Second, the WCEE development algorithm in the analytics subsystem should continuously be enhanced. This is because there is not the perfect algorithm to analyze all data. The algorithm suggested in this dissertation is only a guideline to develop the WCEE. Therefore, the other analytics algorithms should be developed to increase the prediction accuracy. Third, the scenarios should be automatically created by the results of the training and analytics. This does not only help the operators to experience various scenarios but also to generate abundant data without manually implementing scenarios. Fourth, the new methodology should be expanded to each extendable area in detail. The new methodology can be a more powerful tool by materializing the logic reflecting the extendable areas. The limitations and future work are summarized as shown in Figure 65.

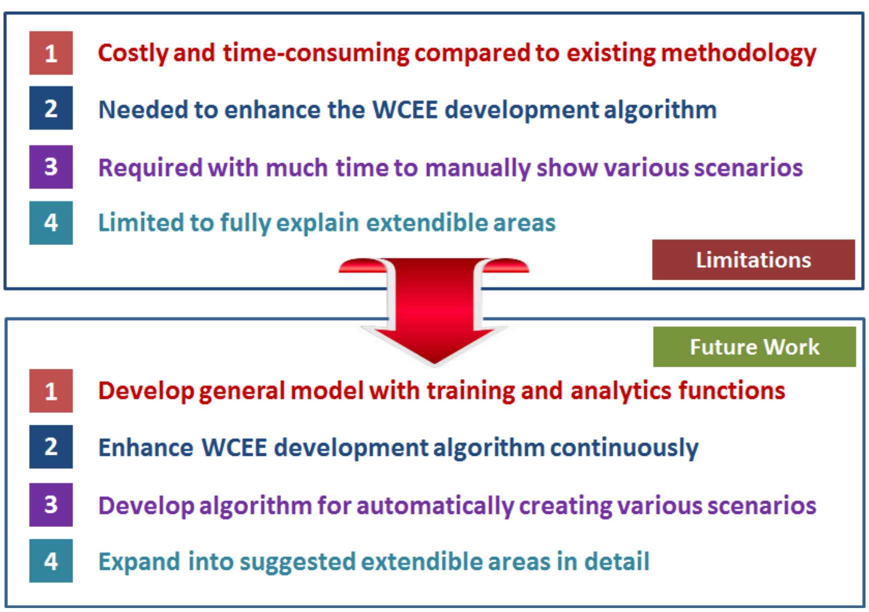

Figure 65. Summary of Limitations and Future Work from This Research

LIST OF REFERENCES

AMSAA. (2017). Army Materiel Systems Analysis Activity. Retrieved from https://www.amsaa.army.mil/About.html

Anderson, C. M. (2004). Generalized Weapon Effectiveness Modeling. In: 2004-06.

Armo, K. R. (2000). *The relationship between a submarine's maximum speed and its evasive capability.* Naver Postgraduate School,

Balci, O. (2007). Verification, Validation, and Testing. In *Handbook of Simulation* (pp. 335-393): John Wiley & Sons, Inc.

Blacklock, J., & Zalcman, L. (2007). *A Royal Australian Air Force, Distributed Simulation, Training and Experimentation, Synthetic Range Environment*. Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Boehm, B. W., Clark, Horowitz, Brown, Reifer, Chulani, . . . Steece, B. (2000). *Software Cost Estimation with Cocomo II with Cdrom*: Prentice Hall PTR.

Boppe, C. W., & Martorella, R. P. (1994). Thrust vectoring/reversing tactics in air-to-air combat. *Journal of Engineering for Gas Turbines and Power*(1), 124.

Calvin, J., Dickens, A., Gaines, B., Metzger, P., Miller, D., & Owen, D. (1993). *The SIMNET virtual world architecture*. Paper presented at the Virtual Reality Annual International Symposium.

Carson, J. S. (1986). Convincing users of model's validity is challenging aspect of modeler's job. *Industrial Engineering*(6), 74.

Chio, T., Weng, Y., & Chang, W. (2010). An algorithm for conducting UAVs' dependability & self-recovery system. *2010 IEEE International Conference on Robotics & Biomimetics (ROBIO)*, 1645.

Chusilp, P., Charubhun, W., & Koanantachai, P. (2014). Monte Carlo Simulations of Weapon Effectiveness Using Pk Matrix and Carleton Damage Function *International Journal of Applied Physics and Mathematics, 4*(4), 280-285.

Connors, C. D. (2015). Agent-based Modeling Methodology for Analyzing Weapons Systems. In: 2015-03-26.

Conte, S. D., Dunsmore, H. E., & Shen, V. Y. (1986). *Software engineering metrics and models*: Benjamin-Cummings Publishing Co., Inc.

Daly, M., & Thorpe, D. (2009). *Balancing Simulated and Live Naval Fleet Training*. Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Davis, P. K. (1992). Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations. In: 1992.

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The mahalanobis distance. *Chemometrics and intelligent laboratory systems, 50*(1), 1-18.

Diebold, F. X. (2012). A personal perspective on the origin(s) and develop-ment of "big data": The phenomenon, the term, and the discipline (ScholarlyPaper No. ID 2202843). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract id=2202843

DoD. (1995). Modeling And Simulation (M&S) Master Plan. In: 1995-10.

Fujimoto, R. M. (1999). *Parallel and distributed simulation.* Paper presented at the 31st conference on Winter simulation.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137-144. doi:10.1016/j.ijinfomgt.2014.10.007

Graebener, R. J., Rafuse, G., Miller, R., & Yao, K.-T. (2004). *Successful joint experimentation starts at the data collection trail-part II.* Paper presented at the Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC '04).

Granowetter, L. (2013). *The WebLVC Protocol: Design and Rationale*. Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Gustavsson, P. M., Björkman, U., & Wemmergård, J. (2009). *LVC aspects and integration of live simulation*. Retrieved from The 2009 Fall Simulation Interoperability Workshop:

Hammes, T. X. (2004). *The sling and the stone : on war in the 21st century*: St. Paul, MN : Zenith Press, c2004.

Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review, 34*(1), 135-174. doi:10.1111/dpr.12142

Hodgson, J. (1957). *A generalized analysis of war gaming (CORG memorandum)*: Tech/Ops, Combat Operations Research Group.

Hopkins, N. J., Marr, H., Stachtchenko, L., & Baker, K. (1966). COST-EFFECTIVENESS OF A SURVEILLANCE DEVICE FOR AN ANTI-AIRCRAFT WEAPON SYSTEM. *CORS Journal, 4*(3), 149-164.

Hsinchun, C., Chiang, R. H. L., & Storey, V. C. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly, 36*(4), 1165-1188.

Hudgins, G. (2007). *The Test and Training Enabling Architecture,TENA, an Important Component in Joint Mission Environment Test Capability (JMETC) Successes.* Paper presented at the International Telemetering Conference Proceedings.

IEEE. (2000). *1516-2000 - IEEE Standard for Modeling and Simulation (M&S) High Level ARchitecture (HLA) - Framework and Rules*. Retrieved from

Jette, B. D. (2000). *Land Warrior*. Paper presented at the Digitization of the Battlespace V and Battlefield Biomedical Technologies II, Orlando, FL, USA.

Ji, B.-L., Liu, Y.-X., Wang, X.-X., & Li, X.-Z. (2005). Algorithm for the combat effectiveness standardization of ground weapons. *Binggong Xuebao/Acta Armamentarii, 26*(1), 64-67.

Jiao, S., Li, W. E. I., Ma, P., & Yang, M. (2013). The Simulation Evaluation System for Weapon Operational Effectiveness based on Knowledge Management. *International Journal of Modeling, Simulation & Scientific Computing, 4*(4), -1. doi:10.1142/S1793962313500207

Jin, C., Pang, Z., Li, H., Yuan, Z., & Wu, J. (2014). *Study on improving operating performance of human factors*. Paper presented at the 13th International Conference on Man-Machine-Environment System Engineering, MMESE 2013, Yantai, China.

Jung, C., & Lee, J. (2010). Combat Effectiveness Based Analysis Methodology for Optimal Requirement of Attack Helicopter Using Simulation. *Journal of the Korea Institute of Military Science and Technology, 13*(6).

Jung, W., Lee, Y.-B., Kim, D.-K., & Kang, S.-J. (2012). A Cost Estimation Development Methodology via CER's Linear Combination. *IE interfaces, 25*(3), 347-356.

Jung, W., Lowe, L., Rabelo, L., Lee, G., & Kwon, O. (2017). A Methodology on Guiding Effectiveness-Focused Training of the Weapon Operator Using Big Data and VC Simulations. *SAE International Journal of Aerospace, 10*(2017-01-2018), 57-67.

Jung, W., Shin, K., Mohamed, A., Rabelo, L., & Lee, G. (2016). Weapon Effectiveness Analysis Using Virtual and Constructive Simulations and Big Data. *Industrial and Systems Engineering Research Conference (ISERC) 2016*.

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, W. K. M., Alam, M., . . . Gani, A. (2014). Big Data: Survey, Technologies, Opportunities, and Challenges. *The Scientific World Journal, 2014*.

Kim, J. (2013). A Study on the Cost-Effec-tiveness Analysis for Survivability Design Opti-mization of Light Armed Helicopter. *Journal ofthe Korea Society for Aeronautical & SpaceSciences*, *11*, 816-865.

Laney, D. (2001). 3-D data management: Controlling data volume, velocityand variety. In *Application Delivery Strategies by META Group Inc*.

Lanman, J., Becker, B., & Samper, W. (2009). *Joint service partnership: extending the live training transformation product line.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference.

Le, Z., Zhong, L., Jianqiang, Z., & Xiongwei, R. (2013). Reduction method of weapon system-of-systems assessment index system based on autoencoder. *Journal of Central South University (Science and Technology), 44*(10), 4130-4137.

Lee, J., Shin, S., Kim, J., Bae, S., & Kim, C. (2015). Interrelation Analysis of UGV Operational Capability and Combat Effectiveness using AnyLogic Simulation. *Journal of Applied Reliability, 15(2)*.

Lee, Y. J., & Lo, C. Y. (1994). Optimizing an Air Defense Evaluation Model Using Inductive Learning. *APPLIED ARTIFICIAL INTELLIGENCE, 8*(4), 645-661.

Li, K., Wang, L., Lv, Y., Gao, P., & Song, T. (2015). Research on the rapid and accurate positioning and orientation approach for land missile-launching vehicle. *Sensors (Basel, Switzerland), 15*(10), 26606-26620. doi:10.3390/s151026606

Li, X., Tan, Y., & Yang, K. (2007). Effectiveness evaluation method for armored weapons SoS based on exploratory analysis. *Systems Engineering and Electronics, 29*(9), 1496-1499.

Liang, K., & Wang, K. (2006). *Using Simulation and Evolutionary Algorithms to Evalute the Design of Mix Strategies of Decoy and Jammers in Anti-torpedo Tactics* Paper presented at the Proceedings of the 2006 Winter Simulation Conference.

Liang, Q., Song, B., & Pan, G. (2006). Fuzzy ideal point method for estimation of life cycle cost and effectiveness of torpedo weapon system. *Binggong Xuebao/Acta Armamentarii, 27*(1), 137-140.

Lim, S.-H., Cho, K.-H., & Park, S. (2009). A Study on the Multi-Criteria Decision Making for Effect Analysis and Decision Making of Weapon System. *Journal of the KIMST, 12*(5), 557-562.

Liou, C.-Y., Cheng, W.-C., Liou, J.-W., & Liou, D.-R. (2014). Autoencoder for words. *Neurocomputing, 139*, 84-96. doi:10.1016/j.neucom.2013.09.055

Liou, C.-Y., Huang, J.-C., & Yang, W.-C. (2008). Modeling word perception using the Elman network. *Neurocomputing, 71*, 3150-3157. doi:10.1016/j.neucom.2008.04.030

Liu, C., Fan, H., Bao, X., & Wang, D. (2013). Design and implementation of air-to-ground weapon launcher detector of a certain airplane. *2013 2nd International Symposium on Instrumentation & Measurement, Sensor Network & Automation (IMSNA)*, 337.

Liu, Y., Zhao, C.-N., Wang, X.-S., Wang, G.-Y., & Feng, D.-J. (2011). Evaluation approach for anti-radiation weapon's effectiveness. *Binggong Xuebao/Acta Armamentarii, 32*(3), 321-326.

Loper, M. L., & Cutts, D. (2010). *Comparative Analysis of Standards Management for LVCAR*. Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Ludvik, F., & Konecny, P. (1996). Influence of the air defence rocket ballistics on the rocket weapon system costs. *Sbornik, 1*, 71-84.

M. J, G. (2005). US begins JHMCS trials in two-seat F/A-18D Hornet. *Jane's International Defense Review, 38*, 25.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *Big Data: The Next Frontier for Innovation, Competition & Productivity*, 1-143.

McDonald, M. L., Upton, S. C., Seymour, C. N., Lucas, T. W., Sanchez, S. M., Sanchez, P. J., . . . Smith, J. R. (2014). *Enhancing the analytic utility of the Synthetic Theater Operations Research Model.* Paper presented at the Proceedings of the 2014 Winter Simulation Conference.

Meng, W. (2012). The new concept of information explosion era - Big Data. *The Merchandise and Quality*.

Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). *Big Data: Prospects and Challenges*.

Noseworthy, J. R. (2008). *The test and training enabling architecture (TENA) supporting the decentralized development of distributed applications and LVC simulations*. Paper presented at the the 12th IEEE/ACM International Symposium on.

Osder, S. (1991). Integrated flight/fire control for attack helicopters. *IEEE/AIAA 10th Digital Avionics Systems Conference*, 287.

Pan, S., Ma, D., & Li, Z. (2006). Queueing network method for the combat effectiveness analysis of mixed-mode air-defense systems. *Binggong Xuebao/Acta Armamentarii, 27*(5), 862-864.

Peng, S.-X., Wang, H.-t., & Zou, Q. (2015). Combat effectiveness evaluation model of submarine-to-air missile weapon system. *Xitong Gongcheng Lilun yu Shijian (Systems Engineering Theory & Practice), 35*(1), 267-272.

PEO-STRI. (2006). *Common Standards, Products, Architectures and/or Repositories (CSPAR) Baseline Document*. Retrieved from U.S. Army Program Executive Office (PEO) Simulation, Training and Instrumentation (STRI):

Perlo-Freeman, S., Fleurant, A., Wezeman, P., & Wezeman, S. (2016). TRENDS IN WORLD MILITARY EXPENDITURE, 2015. *SIPRI Fact Sheet*, 1-8.

Qi, Z., Liu, X., Liang, W., & Niu, Y. (2011). Fuzzy evaluation approach to missile defense effectiveness based on layered weight coefficients. *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice, 31*(9), 1811-1815.

Rabelo, L., Eskandari, H., Shaalan, T., & Helal, M. (2007). Value chain analysis using hybrid simulation and AHP. *International Journal of Production Economics, 105*(2), 536-547. doi:http://dx.doi.org/10.1016/j.ijpe.2006.05.011

Ru, W., & Gao, X. (2012). Weapon Target Allocation Based on Cost-Effectiveness. *Fire Control and Command Control, 37*(2), 57-60.

Ruan, M.-Z., Li, Q.-M., & Liu, T.-H. (2010). Modeling and optimization on fleet antiaircraft firepower allocation. *Binggong Xuebao/Acta Armamentarii, 31*(11), 1525-1529.

Sargent, R. G. (2011). Verification and validation of simulation models. *Proceedings of the 2011 Winter Simulation Conference (WSC)*, 183.

Schmidt, J. W. (1978). *Introduction to simulation modeling.* Paper presented at the 10th Conference: Winter Simulation.

Seo, K., Song, H., Kwon, S., & Kim, T. (2011). Measurement of Effectiveness for an Anti-torpedo Combat System Using a Discrete Event Systems Specification-based Underwater Warfare Simulator. *Journal of Defense Modeling & Simulation, 8*(3), 157.

Sharma, J. K. (2012). *Business Statistics*: Dorling Kindersley.

Shi, Q., & Abdel-Aty, M. (2015). Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C*, 380. doi:10.1016/j.trc.2015.02.022

Skow, A. M. (1992). Agility as a contributor to design balance. *Journal of Aircraft*(1), 34.

Song, H.-L. (2015). Research on penetration and damage abilities of anti-ship missile against warships. *Binggong Xuebao/Acta Armamentarii, 36*, 77-82.

Steinman, J. S. (2013). *The Roadmap*. Retrieved from Simulation Interoperability Workshop:

Steinman, J. S., & Hardy, D. R. (2004). *Evolution of the standard simulation architecture: DTIC Document*. Retrieved from

Tolk, A. (2012). *Engineering principles of combat modeling and distributed simulation. [electronic resource]*: Hoboken : Wiley, c2012.

Tong, Z., Ye, G., Lu, Y., & Hui, X. (2007). Method of evaluation of air-to-air combat effectiveness of simulations operated by pilots. *Journal of System Simulation, 19(8)*, 1680-1682, 1695.

Tuttle, R. (2003). New tactics. *Aviation Week and Space Technology, 158*(23), 50-51.

Wang, L.-y., Dong, Y.-f., Jiang, Y.-y., & Zhang, H.-x. (2007). Synthesized index evaluation model for anti-ship combat capability of attack plane. *Systems Engineering and Electronics, 29*(5), 771-773.

Weatherly, R. M., Wilson, A. L., Canova, B. S., Page, E. H., Zabek, A. A., & Fischer, M. C. (1996). *Advanced distributed simulation through the aggregate level simulation protocol*. Paper presented at the Twenty-Ninth Hawaii International Conference on the System Sciences.

Wilcox, P. A., Burger, A. G., & Hoare, P. (2000). Advanced distributed simulation: a review of developments and their implication for data collection and analysis. *Simulation Practice and Theory, 8*, 201-231. doi:10.1016/S0928-4869(00)00023-9

Wu, R.-C., Zhang, F.-L., Zhang, J.-B., & He, Q. (2014). Application of fuzzy comprehensive evaluation in weapon equipment systems. *Computer Modelling and New Technologies, 18*(5), 138-142.

Yan, D.-W., Gu, L.-X., Guan, Q.-S., & Sun, P. (2007). Combat effectiveness modeling and evaluation of hypersonic cruise missiles. *Binggong Xuebao/Acta Armamentarii, 28*(6), 725-729.

Yin, T., & Xie, W. (2015). Static analysis of acquisition performance of weapon and equipments based on cost and effectiveness. *Journal of Equipment Academy, 26*(5), 36-40.

Yoo, S. K., Lee, J. S., & Baik, D.-K. (2012). A Methodology for Effectiveness Analysis of Future Weapon System Using a PLAF Based Simulation System. *Advanced Methods, Techniques & Applications in Modeling & Simulation*, 336.

Zalcman, L., Blacklock, J., Foster, K., & Lawrie, G. (2011). *An Air Operations Division Live, Virtual, and Constructive (LVC) Corporate Interoperability Standards Development Strategy: DTIC Document*. Retrieved from

Zhao, L., Yin, J., & Song, Y. (2016). An exploration of rumor combating behavior on social media in the context of social crises. *Computers in Human Behavior, 58*, 25-36. doi:10.1016/j.chb.2015.11.054

Zheng, J., Zhang, A., & Wang, Y.-H. (2009). Coordination method of air combat behavior based on Multi-Agent. *Systems Engineering and Electronics, 31*(11), 2663-2667.

Zheng, Y. (2011). Research on Combat Effectiveness Evaluation of Radar EW System Based on Bayesian Network. *Advanced Materials Research, 204*(1), 1697.

Zhou, D., Zhang, K., Zeng, L., & Zhang, K. (2015). Damage assessment of airborne laser weapon to anti-missile. *2015 34th Chinese Control Conference (CCC)*, 223.