

Development of standard loss analysis model using big data: Focusing on Republic of Korea army

Lee, Seungho* · Lim, Changkyu** · Lee, Chahwa*** · Cho, Yongju**** · Kim, Jaeh*****

ABSTRACT

The Republic of Korea Army conducts simulations during peacetime using ground operation resource requirements analysis model (GORRAM) to determine potential losses when at war based on the latest operation plan. Although war-game simulation can yield reliable results, it takes considerable amount of time and effort to build a database and generate scenarios. Therefore, a study is required to supplement the detailed war-game simulation method to quickly determine expected losses. Using data built-in GORRAM, we tested the significance of four factors using beta regression analysis. While multiple regression is most commonly used to model the causality, beta regression is a powerful method for modeling response variables in the (0,1) range, such as the loss ratio. We verified that three factors, namely 'topography', 'operational posture', and 'friend/foe power ratio' were related to loss. This study proposes a new method for calculating the expected loss in real-time, overcoming a limitation of existing war-game simulation methods.

Keywords : war time resource requirement analysis, ground operation resource requirements analysis model (GORRAM), standard loss analysis model, weapon effectiveness index

* (First Author) Center for Army Analysis and Simulation (CAAS), Department of resource analysis, analysis planning officer, <https://orcid.org/0000-0001-9385-167X>

** (Co-Author) CAAS, Army personnel command, personnel officer

*** (Co-Author) CAAS, Department of resource analysis, head of resource analysis, <https://orcid.org/0000-0002-7066-0026>

**** (Co-Author) CAAS, chief of CAAS, <https://orcid.org/0000-0002-4018-6710>

***** (Corresponding Author) Inha University, Dept. of data science, assistant professor, <https://orcid.org/0000-0001-7831-6353>

I. INTRODUCTION

It is important to predict the outcome of a battle in wartime and to prepare resources, such as personnel, equipment, fuel, and ammunition, based on the predicted results. Recently, Ukraine did not fully reserve the artillery ammunition required for the war, therefore they could not respond effectively to Russia's attacks. The Center for Army Analysis and Simulations (CAAS) has been analyzing wartime resource requirements since the 1980s. The analysis of wartime resource requirements is conducted periodically whenever the operation plan and environment are changed. CAAS conducted the analysis using the ground operation resource requirements analysis model (GORRAM) developed in 2010. Using this model, we analyzed the war-time resource requirements of personnel, equipment, fuel, ammunition, repair parts, and materials. The results of wartime resource requirement analysis (WRRRA) are used as the basis for writing various war documents related to inventory requirements and military force, such as the joint strategic objective plan (JSOP) and oil war stockpile documents, according to the directive of the Korea Ministry of Defense. Although war-game simulation can provide relatively reliable results (De Lima Filho et al., 2022; Mittal & Davidson, 2020; Turnitsa, Blais, & Tolk, 2021), it is impossible to provide the results of judging the expected loss after simulating the war-game within a short period of time (e.g., Brathen, Seehuus, & Mevassvik, 2021; Hujer, Kratky, & Farlik, 2020).

During the recent combined command post-training (CCPT), we analyzed the expected loss reflecting the operational environment within a short period of time. This demand is a great challenge. This is because it takes several weeks to simulate a changed operating environment using war-game simulation. Therefore, a study is required to supplement detailed war-game simulation methods and to quickly determine expected losses by reflecting real-time changes in battlefield situations during war and CCPT. The purpose of present study is to develop a standard loss analysis model to overcome the limitations of the existing war game model, using a regression equation to predict the loss of equipment and troops. The three areas of focus in the standard loss analysis model are as follows.

First, we verified which factors were related to the loss of personnel and equipment in previous studies and documents using statistical techniques, such as topography, operational posture, and force ratio. Second, we calculated the regression formula for loss to predict the loss of personnel and equipment based on the war-game simulation data accumulated in GORRAM. Finally, we developed a program that could be used during wartime and CCPT. The statistical

method used in this study was a beta regression model. The beta regression model is a type of generalized linear regression and is a method for exploring the causal relationship between response variables from 0 to 1, such as ratio and interval, and explanatory variables. Considerable number of studies have been published on the application and theory of beta regression since Ferrari and Cribari-Neto (2004) proposed generalized linear models. For instance, standardized residuals of beta (Espinheira, Ferrari, & Cribari-Neto, 2008), improved methodologies (e.g., mixed and longitudinal models) for beta regression (i.e., Figueroa-Zúñiga, Arellano-Valle, & Ferrari, 2013; Hunger, Döring, & Holle, 2012; Ospina & Ferrari, 2012; Schmid et al., 2013; Simas, Barreto-Souza, & Rocha, 2010), and application case for beta regression in the natural sciences (Geissinger et al., 2022).

The remainder of this paper is organized as follows. Section 2 briefly introduces the force scoring mechanisms, including the background of the standard loss analysis model in the ROK Army. In section 3, we provide the estimated regression coefficients using the beta regression method. Section 4 verifies the regression results of the standard loss-analysis model. The development, configuration, and verification of standard loss-analysis model is explained in Section 5. Finally, we conclude this paper with a brief summary and discuss future work in Section 6. The approximate analytical procedures are presented in Table 1.

<Table 1> Analysis procedure

Step	Contents
1	Literature study on loss influencing factors
2	Verification of loss influencing factors by beta-regression
3	Development of the standard loss analysis model
4	Verification of the standard loss analysis model

II. PREVIOUS RESEARCH

Several force scoring mechanisms exist, such as the weapon effectiveness index/weighted unit value (WEI/WUV), division equivalent firepower (DEF), tactical numerical deterministic model (TNMD). For more details, please refer to Allen. The Dupuy Institute developed TNMD to quantify combat outcome trends (Dupuy et al., 2011). TNMD is an empirically based combat model with a database derived from historical research. More than 200 examples of 20th century

combat, mostly World War II, and the 1967 and 1973 Arab-Israeli Wars. They extrapolated historical trends to combat the development and changes in firepower and mobility technologies. This methodology had some limitations, primarily owing to the lack of accounting for situation-dependent combined arm effects. To overcome this limitation, research and development (RAND), which was founded as an American nonprofit global policy think tank in 1948, developed the situational force scoring (SFS) methodology (Allen, 1992). The SFS was developed to improve the representation of ground force close combat and to provide an alternative extrapolation mechanism for use in the more detailed weapon-on-weapon model. The SFS can calculate the advanced and loss rates of personnel and equipment. The SFS consists of three steps: varying the strength of each category of weapons, modifying category multipliers to account for shortages in the combined arms mix, and calculating combat outcomes.

The ROK Army developed a predictive model called the future operation prediction model to calculate the advanced rate and loss rate in 1997. This model was developed based on a TNDM and SFS methodology. None of the previous studies reflected the operating environment of the Korean Peninsula, South Korea's military doctrine, or various weapon systems. These damage assessment methods that simply use the weapon's effectiveness index can be problematic in terms of accuracy. Most importantly, it does not show the factors related to combat loss of personnel and equipment. Here, we verified the factors, namely, topography, operational posture, and others. A regression formula was extracted to predict the loss rate. Finally, we developed a standard loss analysis model to overcome the limitations of the existing war game model.

III. VERIFYING THE FACTORS OF LOSS RATE

3.1 Definition of terms and factors effect

Prior to verifying the loss factors, we define terms related to the loss. The operational posture is the form of operation for the performance of the attacker and the defender's planned missions. The force ratio of friend and foe is the difference in the ratio between the ally's and opponent's military force, based on the power index. The operational tempo is the relative operational speed of the enemy. Here are the loss factors and their effects based on previous research including 'How to make war: A comprehensive guide to modern warfare', etc. (Dunnigan, 2003). Existing studies, such as 'How to make war,' divided the loss influencing factors into operational,

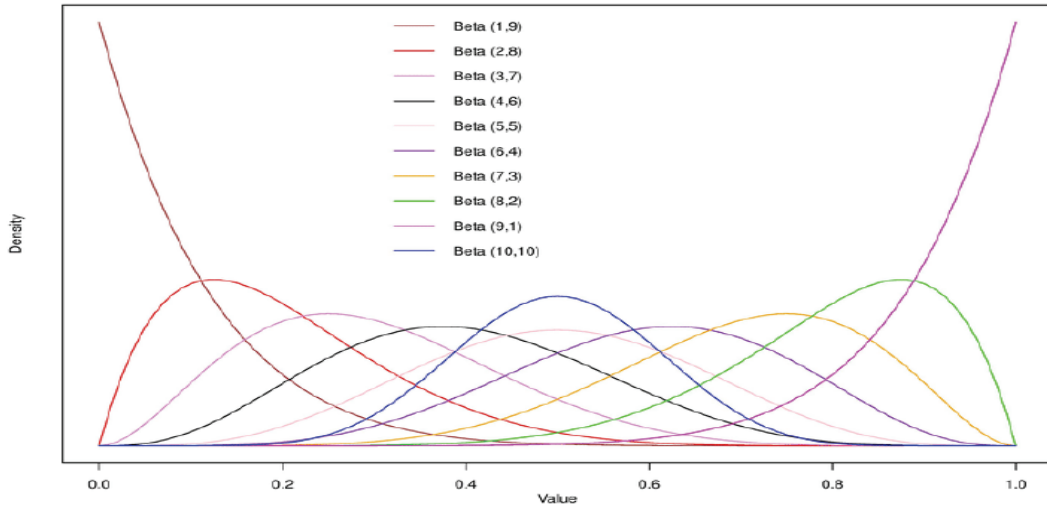
environmental, and intangible factors and presented them, as listed in Table 2. These loss factors, which are listed in Table 2, were verified using statistical methods. For reference, the number in parentheses for each loss factor indicates the degree of impact on the loss.

<Table 2> Loss influencing factors and their effects

Operational factors (effects)	Environment factors (effects)	Intangible power factors
Unit size (50 - 200%)	Terrain (30 - 100%)	Will to fight
Operational posture (30 - 150%)	Weather (30 - 100%)	Discipline
Force ratio (70 - 160%)	Season (65 - 100%)	Level of training
Intensity of operation (70 - 150%)	Time of battle (50 - 100%)	Combat fatigue
Raid achievement (120 - 150%)		

3.2 Overview of Beta Regression

First, as proposed by Ferrari and Cribari-Neto (2004), beta regression is a statistical method in which the dependent variable is based on beta distribution. Beta regression employs a link function to map the data in real space to the bounded interval (0,1) and then performs a regression on the beta distribution using maximum likelihood estimation. While beta regression is most readily applied to the modeling of rates and proportions provided the constrained interval, because of the flexibility of the underlying beta distribution, beta regression is utilized in a wide range of disciplines including medicine, finance and economics, and social science. Because the beta distribution is an extremely flexible distribution, beta regression can be useful for dependent variables $y \in (0, 1)$, as shown in Figure 1. A beta regression model is a maximum likelihood derivation combining link functions, such as logit and probit, with reparametrization of the beta distribution. While multiple linear regression is the most widely used approach in modeling a variable with multiple explanatory variables when the output variable ranges between $(-\infty, \infty)$, beta regression is particularly useful for modeling ratios because the beta distribution is constrained between (0, 1).



<Figure 1> Beta densities for different parameter values

A typical beta probability density distribution is as follows:

$$f(y; p, q) = \frac{\Gamma(P+q)}{\Gamma(P)\Gamma(q)} y^{p-1}(1-y)^{q-1}, 0 < y < 1 \tag{1}$$

where $p, q > 0$ and $\Gamma(\cdot)$ denotes the gamma distribution. These parameters can be transformed to $\mu = p / (p+q)$ and $\phi = p + q$, as follows:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, 0 < y < 1 \tag{2}$$

where $0 < \mu < 1$ and a known precision parameter $\phi > 0$. When $E(y) = \mu$ and $VAR(y) = \mu(1-\mu)/(1+\phi)$, the response variable follows a beta distribution, which is denoted as $y \sim B(\mu, \phi)$. Given random variables y_1, \dots, y_n such that $y_i \sim B(\mu_i, \phi)$, the beta regression is defined as follows:

$$g(\mu_i) = x_i^T \beta = \eta_i \tag{3}$$

where $g(\mu_i): (0,1)$ is a link function that is strictly increasing and twice differentiable; β is a regression vector; x_i is an independent vector; and η_i is a linear predictor. Then,

$$\mu_i = g^{-1}(x_i^T \beta) \tag{4}$$

is estimated using the maximum likelihood. It is assumed that the precision parameters in Equations (2) and (3) are identical. However, this assumption may not be applicable to many practical problems. Simas et al. assumed $y_i \sim B(\mu_i, \phi_i)$ and expanded the regression model as follows [6].

$$g_1(\mu_i) = x_i^T \beta = \eta_{1i}, \quad g_2(\phi_i) = Z_i^T \gamma = \eta_{2i} \tag{5}$$

where β and γ denote regression vectors; x_i and z_i denote independent variable vectors; and η_{1i} and η_{2i} denote linear predictor variable vectors. Such models are referred to as variable dispersion beta regression models.

3.3 Verification of loss factors using statistical method

Among the various loss factors in Table 2, we selected four factors, such as topography, operational posture, force ratio, and operational period, which were accumulated as big-data in GORRAM. We used beta regression to verify these factors. Table 3 lists the verification results.

<Table 3> Verification results

Variables		Estimated value	Probability of significance	Significance
Response variable	Loss rate		$< 2.2 \times 10^{-16}$	
Explanatory variable	Topography	0.04713	2.71×10^{-10}	Significant
	Operational posture	0.18002	8.48×10^{-15}	
	Force ratio	0.54572	$< 2.0 \times 10^{-16}$	
	Operational period	0.01223	0.22626	Insignificant

From the regression output, three of the four loss-influencing factors were observed to be significant. Among the three, the force ratio, difference between the ally and opponent’s military force based on the power index, had the largest influence on the daily loss rate. Operational posture, the form of operations for the planned mission of the attacker and defender, and the effects of topographic characteristics and weather were also observed to be statistically

significant explanatory variables, while their influence on the daily loss rate was much smaller than that of the force ratio. The operation period (number of days) was observed to be statistically insignificant in terms of explaining the loss rate.

IV. CALCULATION OF REGRESSION FORMULA

4.1 Building the database

The simulation output is stored in GORRAM in a tabular format, with the loss of resources indicated by the type of personnel and equipment for each corps/division and day of battle. To construct the database useful for this study, the original data were organized according to the three variables identified as significant in predicting the loss rate, geometry, operational posture, and force ratio. The appropriate classifications for geometry and operational posture listed in Table 4 were applied based on the domain knowledge of the operational condition and function of each corps/division in the ROK Army. In terms of topography, it is divided into flatland, hill, and mountain because the loss is different depending on the place where the war takes places. From the same perspective, the type of operation is divided into attack, defense, fixing, and reserve.

<Table 4> Classification of loss influencing factors

Topography	Operational posture	Force ratio
Flatland	Attack	1:1
Hill	Defense	2:1
Mountain	Fixing	~
Reserve~		

We choose 5 representative weapon systems - mortar, artillery, tank, armored vehicle, communication - in use by the Republic of Korea Army in Table 5. Considering the geopolitical characteristics of the Korean Peninsula, the Republic of Korea Army should carry out both conventional and advanced wars at the same time. It was divided into categories for mortar which are crucial for infantry, artillery equipment, equipment necessary for armoured units, and command and control communication equipment to carry out the fight in this circumstance.

For each of the defined groups, representative equipment was presented.

<Table 5> A detailed list of equipment

Equipment	
Mortar	60 mm, 81 mm, 4.2"
Artillery	KH-179, K-55, K-9, M101A1
Armored vehicle / Tank	K200, K277, K288, K21 / K1, K2
Communication	P-999K, V-946K, P-964K

4.2 Calculating the regression formula

After building the database, we calculated the beta regression formula, as listed in Table 6, using loss-influencing factors, such as topography, operational posture, and force ratio, which were verified in Section 3. The variables in the regression formula are displayed as random numbers for military security reasons. Although not all regression formulas were included in this paper owing to military security concerns, we were able to achieve findings that were supported by activities. All standard losses for the important equipment listed in Table 6 can be predicted. We employed a beta regression model at the time, with the loss rate between 0 and 1.

<Table 6> Regression formula results

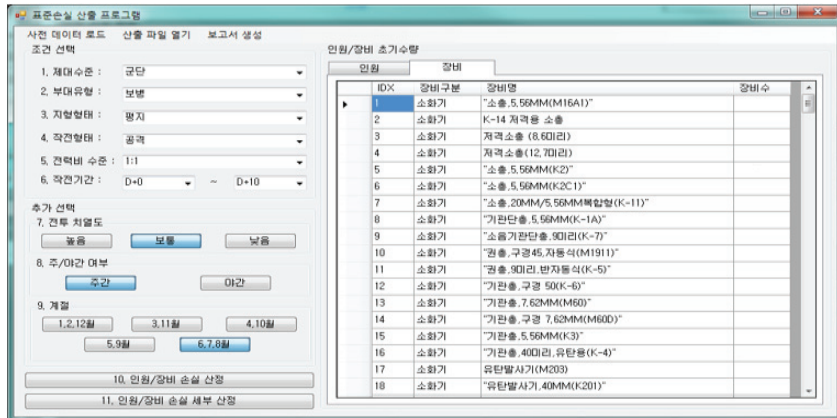
Equipment	a_1	a_2	a_3	a_0
60 mm mortar	0.00075	0.00548	-0.00419	0.00578
K-9 artillery	0.00058	0.00601	-0.00101	0.00441
K1 tank	0.00071	0.00153	-0.00712	0.01021
~	~			

V. Development of Standard Loss Analysis Model

5.1 Development and configuration of the model

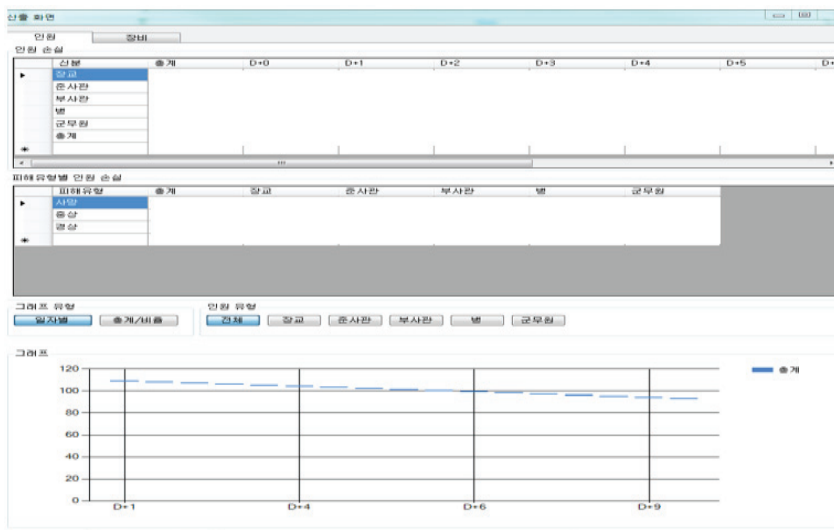
A standard loss analysis model was developed using C#. C# is an object-oriented programming language developed by Microsoft that is intended to combine the computing power of C++ with the programming ease of Visual Basic. We chose the C# language as a development tool, considering processing speed and management after development. Figure 2 shows the standard loss analysis model. To calculate the loss of personnel and equipment, we followed nine steps. The user can choose unit (division or corps) and type of unit (infantry or

artilleryman, etc.), type of topography (flatland, hill, and mountain), type of operational posture (attack, defense et al), power ratio, etc. We can choose additional factors, such as the intensity of the operation, season, and time of operation.



<Figure 2> The input window of the standard loss analysis model

Figure 3 shows the result window in which the loss of personnel is calculated. The model has a multiwindow pop-up function, and it is possible to compare the results of various operational situations in a very short time. This allows for a more precise prediction of the loss of various types of war participants, such as officers and men. We can assist prepare for war effectively by predicting losses from D-day to D+60 days. These numbers were blanked out for military security concerns.



<Figure 3> The result window of the standard loss analysis model

VI. SUMMARY AND FUTURE WORKS

To overcome the limitations of existing war-game simulations, the CAAS has conducted research on the standard loss analysis model. We verified the loss factors, such as topography, operational posture, and force ratio, using the simulation data of GORRAM using beta regression. Based on the verification, we calculated a regression formula to predict the loss of personnel and equipment. Finally, we developed a standard loss analysis model using C# and verified the model by comparing the results of the model with the existing studies. Throughout this study, we analyzed the war-time resource requirements within several hours by comparing 1 - 2 weeks of the existing war-game model. Through this study, we were able to overcome the time limitation of the existing war-game model and improve the accuracy of the method by simply evaluating loss with the index of the weapon system. In addition, the output of this study will be used as a means of judging the expected loss by reflecting the changed operational environment from this joint US - ROK exercise. This model will be distributed to the Corps and Division for analyzing the operational plan by comparing the loss of various operational plans. Finally, we must update the regression formula using more recent simulation data. In addition, we will expand the variety of target equipment from 10 - 20 types to 100 - 120 types. After improving the standard loss analysis model by the end of this year, it will be used on a trial basis in the next ROK-US CCPT.

Acknowledgements

We would like to thank Editage (www.editage.co.kr) for the English language editing.

Declaration of Conflicting Interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author contributions

Conceptualization: LS, LC, and CY; Resources and Data curation: KJ; Investigation: LS and KJ; Methodology: LS and KJ; Writing (Original Draft): LS; Writing (Review and Editing): LS and KJ; Project administration and Supervision: LS, LC, and CY.

Reference

- Allen, P. D. (1992). Situational for scoring: Accounting for combined arms effect in aggregate combat models (A Rand note). RAND CORP SANTA MONICA CA. <https://apps.dtic.mil/sti/citations/ADA260635>
- Brathen, K., Seehuus, R. A., & Mevassvik, O. M. (2021). *Simulation Support to Wargaming for Tactical Operations Planning*. *Simulation and Wargaming*, 223–247. <https://doi.org/10.1002/9781119604815.ch9>
- De Lima Filho, G. M., Kuroswiski, A. R., Medeiros, F. L. L., Voskuijl, M., Monsuur, H., & Passaro, A. (2022). Optimization of Unmanned Air Vehicle Tactical Formation in War Games. *IEEE Access*, 10, 21727–21741. <https://doi.org/10.1109/ACCESS.2022.3152768>
- Dunnigan, J. F. (2003). *How to make war: A comprehensive guide to modern warfare in the twenty-first century*. Harper Collins. https://books.google.co.kr/books/about/How_to_Make_War_Fourth_Edition.html?id=ULeyNtKzLT8C&redir_esc=y
- Dupuy, A., Ishutkin, V., Pickl, S., & Tschiedel, R. (2011). Judgment based analysis via an interactive learning software for modern operations research education and training. In *Operations Research Proceedings 2010* (pp. 623–628). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20009-0_98
- Espinheira, P. L., Ferrari, S. L., & Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, 35(4), 407–419. <https://doi.org/10.1080/02664760701834931>
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815. <https://doi.org/10.1080/0266476042000214501>
- Figueroa-Zúñiga, J. I., Arellano-Valle, R. B., & Ferrari, S. L. (2013). Mixed beta regression: A Bayesian perspective. *Computational Statistics & Data Analysis*, 61, 137–147. <https://doi.org/10.1016/j.csda.2012.12.002>
- Geissinger, E. A., Khoo, C. L., Richmond, I. C., Faulkner, S. J., & Schneider, D. C. (2022). A case for beta regression in the natural sciences. *Ecosphere* 13(2), e3940. <https://doi.org/10.1002/ecs2.3940>
- Hujer, V., Kratky, M., & Farlik, J. (2020). Utilization of Modeling and Simulation in the Design of Air Defense. In: Mazal, J., Fagiolini, A., Vasik, P. (eds) *Modelling and Simulation for Autonomous Systems*. MESAS 2019. Lecture Notes in Computer Science, 11995. Springer, Cham. https://doi.org/10.1007/978-3-030-43890-6_23

- Hunger, M., Döring, A., & Holle, R. (2012). Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC Medical Research Methodology*, *12*(1), 1-12. <https://doi.org/10.1186/1471-2288-12-144>
- Mittal, V., & Davidson, A. (2020). Combining wargaming with modeling and simulation to project future military technology requirements. *IEEE Transactions on Engineering Management*, *68*(4), 1195-1207. <https://doi.org/10.1109/TEM.2020.3017459>
- Ospina, R., & Ferrari, S. L. (2012). A general class of zero or one inflated beta regression models. *Computational Statistics and Data Analysis*, *56*(6), 1609-1623. <https://doi.org/10.1016/j.csda.2011.10.005>
- Schmid, M., Wickler, F., Maloney, K. O., Mitchell, R., Fenske, N., & Mayr, A. (2013). Boosted - beta regression. *PloS one*, *8*(4), e61623. <https://doi.org/10.1371/journal.pone.0061623>
- Simas, A. B., Barreto-Souza, W., & Rocha, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* *54*(2), 348-366. <https://doi.org/10.1016/j.csda.2009.08.017>
- Turnitsa, C., Blais, C., & Tolk, A. (Eds.). (2021). *Simulation and Wargaming*. Wiley. Retrieved from <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119604815>

Received: 2 July, 2022.

Accepted: 10 August 2022

Published: 31 August 2022