Techniques to assess the quality of armor performance measurement

D. Bourget, M. Bolduc

Defence Research and Development Canada – Valcartier Research Center, 2459 Pie-XI Blvd North
Québec (Québec) G3J 1X5, Canada, Daniel.Bourget@drdc-rddc.gc.ca

Abstract: Ballistic limit and proofing velocity tests are used for decades to assess the performance of armors in general. Many papers discussing the subject have been presented in the preceding PASS conferences concerning both: methods to efficiently execute testing and methods to calculate the ballistic limit and proofing velocity. This paper discusses about the later. Recent important changes in this field include the introduction of fitting techniques to calculate the ballistic limit using either Logit (NIJ 0101.06) or Probit (NATO STANAG 2920) link functions and the calculation of the uncertainties of the ballistic limit values. Different link functions and techniques to assess their Goodness of Fit towards the experimental data are discussed in this paper. In particular, the Wald test, the Likelihood Ratio test, the Anderson-Darling test, the Sensitivity/Specificity Analysis and the Information Criteria are discussed.

1. INTRODUCTION

Armour performance, against specific threats can be evaluated using ballistic limit (BL or V_{50}) and proofing velocity (V_{proof}) tests. While the V_{proof} test is used to assess if the armour actually protects against a threat and does not consider the residual capacity of the armour, the BL test enables the assessment of the absolute resistance of an armour against a threat and therefore, enables the ranking of armour performance. For this reason BL tests are specified in most ballistic standard along with V_{proof} tests. Up until recently, BL value was calculated based on the simple average of the impact velocity of 3 (or 5) partial and 3 (or 5) complete perforations of the armour within a specific velocity spread. Discussions within the different Standard committees showed that this simple way of calculating the BL value does not provide the full picture of the armour capability and that more information can be drawn out of the raw data. Amongst them, the estimation of the impact velocity corresponding to low probability of perforation (e.g. V_{05} for 5% probability of penetration and V_{01} for 1% probability of penetration) is of interest as these values are highly related to the protection capability of the armour and the V_{proof} test results. In addition, the uncertainty related to the V_{50}, V_{05} and V_{01} estimations are needed for further armour performance comparison purposes. To assess the parameters mentioned above, techniques consisting of fitting probabilistic curves through the pass/fail BL test data point were put forward and consequently, made their way to ballistic standards. Not a simple calculation anymore, BL test data exploitation necessitates techniques that need to be programmed to enable a robust and trustworthy evaluation. In close relation with the estimate of the V_{50}, V_{05}, V_{01} values and their uncertainties, the statistical significance and the goodness of fit of the model to estimate these values need to be assessed. This enables the analyst to determine how well the fitted model used in the fitting process corresponds to the reality of the experimental data.

To fulfill these needs, the BLC (Ballistic Limit Calculator), a MS Excel VBA program was created to:

a. Calculate the BL and V_{proof} related values,
b. Calculate the errors on these values, and
c. Assess the goodness of fit of the model, be it Logit, Probit or others, to the experimental data.

This paper describes the statistical aspect of the BLC and the different probabilistic models that are used within the software. It also provides examples of how to use the software for ballistic data analysis.

2. BALLISTIC AND STATISTICAL BACKGROUND

2.1 General

The Ballistic Limit is the velocity at which the probability of perforation of an armour is 50% and it is defined for a specific armour against a specific threat. It is calculated based on a series of shots fired at an armour following a pre-defined impact pattern for a range of impact velocity above and below the estimated BL value. The sequence of impact velocity follows the up and down procedure. It is well described in NATO STANAG 2920 (Edition 3) [1] and [2] and in the NIJ 0101.06 [3] documents. The

https://doi.org/10.52202/078352-0035

up and down procedure is used to minimize the number of shots necessary to determine the 50% probability value. For each firing, the dichotomous impact status (either a perforation (1) or a non-perforation (0)) is recorded.

Using the impact velocity as independent variable and the impact status as the dependent variable, the probability of perforation versus impact velocity is modelled using specific probabilistic “link functions” for dichotomous data. Amongst the large number of possible link function, the Probit and Logit cumulative distributions are the most often specified, [1] to [3]. The model parameters are defined using a curve fitting technique that uses Maximum Likelihood estimation (ML). Both, NIJ and NATO STANAG suggest using ML estimation and have described the general procedure ([1] and [3]). For research purposes the current version of BLC includes 3 other link functions: the Gompit distribution, the Scobit distribution and the Weibull distribution. As part of the BL and its variance estimation, it is necessary to assess how well the data are modelled by the different model functions. This assessment can be done using typical goodness-of-fit tests available for dichotomous dependent variables.

2.2 Link functions

The Probit link function, also called the cumulative normal distribution (Φ), has a mean μ and a variance σ^2. The Logit link function also has two parameters, α and β. Both link function probability of perforation distribution (Pp) equation for an impact velocity (V_S) can be found in Table **1**, 2^{nd} column. Those link functions are symmetrical around their means (i.e their skewness is null) and they have very similar shape as can be seen in Figure 1. The Probit link density probability function is steeper in the middle and more quickly approaches zero on the left and on the right compared to the Logit link function [4].

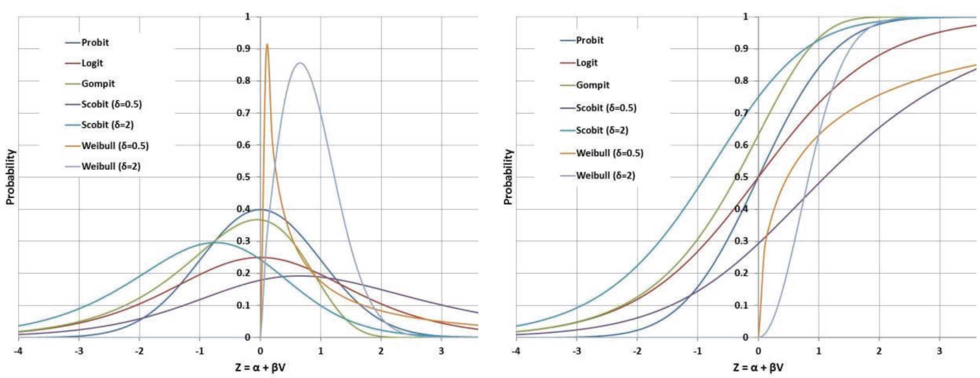

Figure 1 – Density probability (left) and cumulative probability (right) distribution curves for the Probit, Logit, Gompit, Scobit and Weibull distributions. For the Scobit and the Weibull distributions, curves for 2 different shape parameter values are shown (δ = 0.5 and δ = 2.0). Notice the difference in the slope around Z = 0 for the different cumulative probability distributions.

The Gompit link function is also called the cumulative complementary log-log (CLogLog) distribution. It has two parameters: α is the location parameter and β is the scale parameter (Table **1**). That link function is not symmetrical and has a fixed negative skewness of -1.1396 (skewed to the left). This means that its cumulative probability function is S-shaped and that it approaches 0 slowly but approaches 1 much more sharply. The Scobit link function [4] is also called the Burr-10 distribution or a Type I skewed-logit function. It has tree parameters: α is the location parameter, β is the scale parameter and δ is the shape parameter (Table **1**). The Logit distribution is a special case of the Scobit distribution: for δ = 1, the Scobit and Logit distributions are equivalent. The Scobit link function is not symmetrical, except when δ = 1. Contrary to the Probit, Logit and Gompit distributions, it does not have limitations in the amount of skewness it can model. The Weibull link function has tree parameters: α is the location parameter, β is the scale parameter and δ is the shape parameter (Table **1**). Like for the Scobit distribution, the Weibull distribution is not symmetric and does not have a fixed skewness value. The typical curve shapes for the different link functions presented above are shown

 https://doi.org/10.52202/078352-0035

in Figure 1. The Scobit and Weibull distributions are presenting a variety of shapes and skewness depending on the shape parameter (δ) value. As concluded by Mauchant in [6] relative to the analysis of ballistic data using the Probit, Logit and Gompit link functions: "*This work shows that the choice of link function, between the Logit, the Probit and the complementary log-log link functions, is not the most important issue in V_{50} ballistic limit performance estimation, since the different GLMs (Generalized Linear Model) examined all gave similar results*". This might have been the case because the Probit, Logit and Gompit link functions have a constant shape parameter and a fixed skewness, which is not the case for the Weibull and the Scobit link functions.

Table **1** presents a summary of the information related to the different link functions presented above. For all link functions, the standard deviation σ of the distribution is equal to $1/\hat{\beta}$. The V_{50} value can be calculated directly from the estimated α, β and δ values following the formulas shown in the last column.

Table 1 – Summary of link function characteristics and V_{50} formula

Link function	Pp (at velocity V_S)	Pp scaling $(\hat{\alpha}+\hat{\beta}V_S) =$	V_S (for probability Pp)	V_{50} value
Probit	$\frac{1}{\sqrt{2\pi}/\hat{\beta}}\int_{-\infty}^{V_S} e^{-\frac{1}{2}(\hat{\alpha}+\hat{\beta}x)^2}dx$	$\Phi^{-1}(Pp)$	$(\Phi^{-1}(Pp)-\hat{\alpha})/\hat{\beta}$	$-\hat{\alpha}/\hat{\beta}$
Logit	$\left[1+e^{-(\hat{\alpha}+\hat{\beta}V_S)}\right]^{-1}$	$-\ln\left(\frac{1-Pp}{Pp}\right)$	$\left[-\ln\left(\frac{1-Pp}{Pp}\right)-\hat{\alpha}\right]/\hat{\beta}$	$-\hat{\alpha}/\hat{\beta}$
Gompit	$1-e^{-e^{(\hat{\alpha}+\hat{\beta}V_S)}}$	$\ln(-\ln(1-Pp))$	$[\ln(-\ln(1-Pp))-\hat{\alpha}]/\hat{\beta}$	$(-0.3665-\hat{\alpha})/\hat{\beta}$
Scobit	$\left[1+e^{-(\hat{\alpha}+\hat{\beta}V_S)}\right]^{-\hat{\delta}}$	$-\ln\left[\left(\frac{1}{Pp}\right)^{1/\hat{\delta}}-1\right]$	$-\left(\ln\left[\left(\frac{1}{Pp}\right)^{1/\hat{\delta}}-1\right]+\hat{\alpha}\right)/$	$-\left(\ln\left[(2)^{1/\hat{\delta}}-1\right]+\hat{\alpha}\right)/$
Weibull	$1-e^{-(\lvert\hat{\alpha}+\hat{\beta}V_S\rvert)^{\hat{\delta}}}$	$[-\ln(1-Pp)]^{1/\hat{\delta}}$	$\left([-\ln(1-Pp)]^{1/\hat{\delta}}-\hat{\alpha}\right)/\hat{\beta}$	$\left([0.6931]^{1/\hat{\delta}}-\hat{\alpha}\right)/\hat{\beta}$

2.3 Curve fitting process

The fitting process consists of finding the estimates of the different distribution parameters (α, β and δ) that maximise the likelihood that the fitted curve corresponds to the experimental data (Maximum Likelihood Estimation, ML). For curve fitting purpose, the ML techniques should be equivalent to the Minimum Least Square (LS) method if the error distribution between the regressed curve and the experimental data points follows a normal distribution. For dichotomous dependent data, the error distribution is not normal and therefore both techniques are not equivalent, ML technique being optimal. The search for the optimal distribution parameters is done using an optimisation algorithms: the Downhill Simplex algorithm. It was published by Nelder and Mead [7] and uses a single-objective optimization approach (in our case, maximising likelihood) for searching the space of p-dimensional real vectors. The p dimensions correspond to the number of parameters in the link function. It only uses the values of the objective functions without any derivative information and therefore, it falls into the general class of direct search methods. It is therefore suitable for problems with non-smooth functions and discontinuous functions which occur frequently in statistics and experimental mathematics.

2.3.1 Maximum likelihood estimation

Briefly, the log-likelihood equation (LL) is the log of the likelihood equation (L) for M Bernoulli trials (dichotomous, i.e. 0 or 1 response, value 1 representing a success) with success probability $P_P(V_S)$. It can be written as follows:

https://doi.org/10.52202/078352-0035

$$LL\,(X|\alpha,\beta,\delta) = Ln\big(L\,(X|\alpha,\beta,\delta)\big)$$
$$= \sum_{i=1}^{M} Resp_i\, Ln\big[P_P(V_{S_i})\big] + (1 - Resp_i)\, Ln\big[1 - P_P(V_{S_i})\big] \tag{1}$$

Where X are the experimental data points composed of $Resp_i$ (the perforation status, either 0 or 1) and the impact velocity V_{S_i}. Note that $P_P(V_s)$ in equation (1) is any of the link functions presented above.

The desirable properties of consistency, normality and efficiency of the ML estimators are asymptotic, i.e. these properties have been proven to hold as the sample size approaches infinity. This has consequences when considering small samples [8]:

i. The small sample behavior of ML estimators is largely unknown. There are no hard and fast rules for selecting sample size. It is risky to use ML with samples smaller than 100, while samples over 500 seem adequate.
ii. While the standard advice is that with small samples larger p-values should be accepted as evidence against the null hypothesis, given that the degree to which ML estimates are normally distributed in small samples is unknown, it is more reasonable to require smaller p-values in small samples.

3. TESTS TO ASSESS GOODNESS-OF-FIT OF MODELS

To assess the Goodness-of-fit of the models is at least as important as fitting the model itself as it provides an indication of how well the model describes the data. Some of the test presented below require an estimation of the Standard Error (SE) of the model parameters. Estimation of the SE of the parameters is also important for the determination of the error bounds of the different parameters. The SE of the parameters are never mentioned when BL values are published, although they should be, at least for quick Goodness-of-fit assessment purpose and error estimation purposes.

3.1 Standard error of the model parameters

Standard errors (SE) of the model parameters are calculated based on the estimated variance (Var, with $SE = \sqrt{Var}$) and covariance of the fitted parameters. The variance of a model with N parameters (β_i, i = 1 to N, in our case N = 3, α, β and δ) is given by equation (2). In BLC, once the best fit parameters for a link functions are defined using the ML technique, a numerical double derivative of the Log-Likelihood function (LL) is calculated for each parameters to define the variance/covariance matrix. For the V_{50} and σ values, which are function of α, β and δ as shown in
Table **1**, the SE is calculated based on the arithmetic of error propagation described in [9], Section 1.4 as well as [10], [11] and [12].

$$\begin{bmatrix} Var_{11} & Var_{21} & Var_{31} \\ \dots & Var_{22} & Var_{32} \\ \dots & \dots & Var_{33} \end{bmatrix} = \begin{bmatrix} -\frac{\partial LL^2}{\partial\beta_1\,\partial\beta_1} & -\frac{\partial LL^2}{\partial\beta_2\,\partial\beta_1} & -\frac{\partial LL^2}{\partial\beta_3\,\partial\beta_1} \\ \dots & -\frac{\partial LL^2}{\partial\beta_2\,\partial\beta_2} & -\frac{\partial LL^2}{\partial\beta_3\,\partial\beta_2} \\ \dots & \dots & -\frac{\partial LL^2}{\partial\beta_3\,\partial\beta_3} \end{bmatrix}^{-1} \tag{2}$$

3.2 The Wald test

The Wald test is obtained by comparing the ML estimate of the parameter (say $\hat{\beta}$) to its standard error (SE($\hat{\beta}$)). The resulting ratio (W) follows a normal distribution under the assumption that β = 0 [13].

$$W = \frac{\hat{\beta}}{SE(\hat{\beta})},\ p-value = P(|z| > W). \tag{3}$$

If p-value ≤ α_{lim} then we can reject the null hypothesis that $\hat{\beta} = 0$. The Wald test was found to behave in an aberrant manner often failing to reject the null hypothesis when the test was significant. It is therefore a conservative test to assess if the parameters are significant or not.

3.3 Likelihood Ratio test

 https://doi.org/10.52202/078352-0035

This test consist of comparing the model likelihood with all its parameters ($\widehat{\alpha}, \widehat{\beta}$ and $\widehat{\delta}$) to the model without one of its parameters ($\widehat{\beta}$). For the purpose of comparing the models with and without the independent variable, the values of the Likelihood functions are compared using equation (4).

$$G = -2\ln\left[\frac{L\left(X|\widehat{\alpha},\widehat{\delta}\right)}{L\left(X|\widehat{\alpha},\widehat{\beta},\widehat{\delta}\right)}\right] \qquad (4)$$

The test statistic G is called the Likelihood Ratio. Under the hypothesis that $\widehat{\beta}$ = 0, the G statistic asymptotically follows a χ^2 distribution with 1 degree of freedom. When the sample size is small, the χ^2 distribution has to be corrected. Many different corrections exists, and reference [14] examined 5 of them in terms for Type I and Type II errors. The authors recommends the Bartlett correction. It is implemented in BLC and its details can be found in [15].

3.4 Anderson-Darling Test

The Anderson-Darling (AD) Test [16] enables the comparison between an assumed cumulative distribution function (a link function with its fitted parameters, F(x,θ)) and the empirical distribution function (EDF, $F_n(x)$) based on the experimental data points. θ is a vector of the parameters ($\widehat{\alpha}, \widehat{\beta}$ and $\widehat{\delta}$). F(x,θ) is compared to $F_n(x)$ using the AD statistic [16]. The AD test can be used to detect the goodness-of-fit for the Probit, Logit, Weibull and exponential distributions. Due to its weight function, the AD test usually makes a more powerful test statistic by emphasizing the tail differences between the empirical distribution function and the assumed cumulative distribution function.

References [17] and [18] compared the power of different tests to the AD tests for small and large samples. They found that: a) the AD test results in excellent power levels compared to the other tests or it is marginally close to the best test, b) sample sizes above 50 can reach power levels above 80% and c) the increase of the allowable Type I error level results in higher power level for the same sample size. It is therefore advisable to accept higher Type I error (15 to 20%) in order to reach acceptable Type II error rates for the typical types of sample size used in ballistic studies.

3.5 Sensitivity, Specificity and Area Under the Curve (AUC)

The use of Sensitivity, Specificity and AUC tries to answer the following question: How good a job the model does of predicting outcomes? Or said another way: What percent of the observations the model correctly predicts? Can my model discriminates between positive and negative outcomes (between complete and partial perforations)?

These questions can be answered by measuring how good the model is to measure true positive and true negative answers while minimizing the number of false positive and false negative. All of these are calculated by comparing the model outcome to the sample measured outcome. First, here are some definitions:

a) **Sensitivity** (or true positive rate) is the proportion of positives (complete perforation) correctly identified by the model. This value has to be has high as possible.
b) **Specificity** (or true negative rate) is the proportion of negative (partial perforation) correctly identified by the model. This value has to be has high as possible. On contrary the value of 1-Specificity is the false negative rate, and it has to be as small as possible
c) **ROC**: Receivers Operating Characteristics. It is a plot of the model sensitivity versus the model false negative rate (1-specificity)
d) The **AUC** is a measure of how well a variable (in our case, impact velocity) can distinguish between two diagnostic groups (complete perforation/partial perforation). AUC is the area under the ROC curve.

Without getting into the details of how the sensitivity, specificity and AUC are calculated we can say that the AUC is at minimum 0.50 (model equivalent to chance) and at maximum 1.0. The higher the AUC between those two limits, the better the model is in predicting the outcome. Theoretically, the AUC is the probability that a randomly selected complete perforation velocity has a higher value (higher impact velocity) than a randomly selected partial perforation velocity.

3.6 The Information Criterion

https://doi.org/10.52202/078352-0035

The Log-Likelihood Ratio cannot be used to compare non nested models. In that case, the Information Criterion (IC) has to be used [19]. It is based on the information theory using the Kullback-Leibler information equation that measures the 'information' lost when approximating reality using a function. The IC accounts for how well the model fits the data and also accounts for the complexity of the model. Model complexity is its ability to fit any data set and can be approximated as the number of parameters in the model [19]. It is well known that some models can be so complex that they can fit any data set [19]. Details on the development of the IC can be found in [20], [21], [22] and [23]. The IC, called the AIC (Akaike Information Criterion [23]) can be calculated using the following: $\mathrm{AIC} = -2\mathrm{LL}(x|\hat{\theta}) + 2p$.

Variable p is the number of parameters in the model and $\mathrm{LL}(x|\hat{\theta})$ is the Log-Likelihood value of the model (equation (1)) for the p estimated parameters ($\hat{\theta}$). For comparison between models, the AIC value as to be as small as possible. AIC is true asymptotically, i.e. when the number of data point is large. The rule of thumb from [19] is that for cases where sample sizes are small ($n < 100$) or where the number of free parameters is large ($p > 5$), the AIC value is biased and therefore it has to be corrected. For the for Logit and Probit link functions, a small sample empirical estimation is used [24]. For the Gompit, Scobit and Weibull link functions the Bootstrap method is used (section 8.3 of [21]).

4. HOW TO USE ALL THESE STATISTICS? (SO WHAT?)

The various outputs of BLC can be used to calculate the following:

1. Statistics for a sample (mean, standard deviation, median, 1st quartile, 3rd quartile, zone of mixed results, skewness, kurtosis, lowest complete and highest partial perforation)
2. Values and uncertainties of the V_{50} and the standard deviation (σ) of an armour for a specific confidence level
3. Based on the calculated uncertainty on the V_{50} and σ values, should more shots be done?
4. Extreme values of V_{50} and σ related to the NATO STANAG 2920 and the NIJ 0101.06 procedures as well as other error distributions
5. Values and uncertainties related to any probability of perforation (e.g. V_{01}, V_{05}, V_{95}, V_{99}) as well as for the V_{proof} value based on normal error and binomial error distributions
6. Identification of the model that most accurately represents the experimental data or: Is the selected model valid?

4.1 Sample statistics

Sample statistics are found under the heading “Sample Statistics” in BLC (Figure 2, part A) and in the “Box Plot” page (Figure 3).

 https://doi.org/10.52202/078352-0035

		SAMPLE STATISTICS					
A	Sample median (m/s) =	513.53					
	Sample mean (m/s) =	513.53					
	Sample σ (m/s) =	17.13					
	Spread (m/s) =	84.00					
	ZMR (m/s) =	16.60					
	LC (m/s) =	505.30					
	HP (m/s) =	521.90					
	Skewness =	-0.32					
	Kurtosis =	2.86					
	Number of data points =	20					
	MONOBIT test p-value for sample randomness =	Cannot reject that data sample is random (p = 0.824)					
	RUN test p-value for sample randomness =	Cannot reject that data sample is random (p = 1.185)					
	Cumulative Sum test p-value for sample randomness =	Cannot reject that data sample is random (p = 0.918)					
		CURVE FIT STATISTICS for CI = 95%					
		Logit (NIJ 0101.06)		Probit (STANAG 2920)		Scobit	
B	V_{50} ± SE (m/s) =	515.25	± 3.8825	515.20	± 3.8295	517.24	± 0.7923
	σ ± SE (m/s) =	6.6496	± 3.3908	10.9791	± 5.3795	0.0841	± 0.0001
	α ± SE (m/s) =	-77.4866	± 24.5426	-46.9257	± 19.3767	-6260.7103	± 9.4536
	β ± SE (m/s) =	0.1504	± 0.0478	0.0911	± 0.0378	11.8953	± 0.0179
	δ ± SE (m/s) =					0.0064	± 0.0022
	Standard Error of Estimate at V_{50} =	0.4205		0.4202		0.4309	
E	Wald test p-value on α =	Reject that α is null (p = 0.001)		Reject that α is null (p = 0.008)		Reject that α is null (p = 0.000)	
	Wald test p-value on β =	Reject that β is null (p = 0.001)		Reject that β is null (p = 0.008)		Reject that β is null (p = 0.000)	
	Wald test p-value on δ =					Reject that δ is null (p = 0.002)	
	Likelihood Ratio test p-value =	Reject that α and β are null (p = 0.005)		Reject that α and β are null (p = 0.005)		Reject that α, β and δ are null (p = 0.004)	
	Osius-Rojek Goodness-of-Fit test p-value =						
	Anderson-Darling Goodness-of-Fit test p-value =	Cannot reject that distribution is Logit (p = 0.382)		Cannot reject that distribution is Probit (p = 0.222)		Reject that distribution is Scobit (p = 0.000)	
	Stukel test for fit of distribution tails p-value =	Logit model tails do not fit the data (p = 0.001)					
	ROC AUC =	0.8586		0.8586		0.8586	
	Model Sensitivity (Proportion of complete perforation correctly identified by the model)	0.7778		0.7778		0.7778	
	Model Specificity (Proportion of partial perforation correctly identified by the model)	0.9091		0.9091		0.9091	
	Threshold probability =	0.5100		0.5100		0.4400	
	Deviance =	18.7279		18.6390		18.2416	
	Pearson X^2 =	16.3956		16.2809		15.8247	
	Nagelkerke pseudo R^2 =	0.4761		0.4799		0.4968	
	Tjur pseudo R^2 =	0.3636		0.3669		0.3830	
	Log Likelihood =	-9.364		-9.320		-9.121	
	Least square =	1.7683E-01		1.7656E-01		1.8566E-01	
	Small Sample Corrected Information Criterion Bias =	2.409		2.223		2.932	
	Small Sample Corrected IC ± SD (m/s) =	23.5451	± 0.0000	23.0840	± 0.0000	24.1046	± 2.6762
C	V_{50} (m/s) CI length from STANAG 2920 method =	15.06		14.46		3.74	
	V_{50} max (m/s) from STANAG 2920 method =	523.04		522.54		517.20	
	V_{50} min (m/s) from STANAG 2920 method =	507.98		508.09		513.45	
	Sigma max (m/s) from STANAG 2920 method =	50.97		83.44		0.20	
	Sigma min (m/s) from STANAG 2920 method =	2.07		3.99		0.08	
	V_{50} (m/s) CI length from Normal CI curves =	19.29		26.92		7.78	
	V_{50} max (m/s) from Normal CI curves =	526.13		532.95		534.10	
	V_{50} min (m/s) from Normal CI curves =	506.83		506.03		-	
		VPROOF STATISTICS					
		$V_{5\%}$ and CI = 95% Normal error distribution					
		Logit		Probit		Scobit	
D	Vproof (m/s) =	495.67		497.14		487.07	
	Max prob of perforation (%) =	27.85		38.24		37.49	
	Min prob of perforation (%) =	0.72		0.14		0.67	
	Vproof max value (m/s) =	504.76		506.21		2199.77	
	Vproof min value (m/s) =	-		-		502.90	

Figure 2 – Example of BLC output

4.2 Values and uncertainties of V_{50} and σ (are more shots necessary?)

Those statistics can be found for each model under the heading "Curve Fit Statistics", Figure 2, part B (Weibull and Gompit models are omitted for clarity). Clearly, the 3 models show similar V_{50} values while the value of σ varies significantly and the standard error also varies significantly. This difference is due to the varying shape and slope of the link functions as seen above (Figure 1). If unsatisfied with the V_{50} standard error obtained, another shot can be done and statistics recalculated to provide another estimate of V_{50} and σ with their SE.

4.3 Extreme values of V_{50} and σ

V_{50} and σ values as calculated using the NATO STANAG 2920 procedure can be found under the heading "Curve Fit Statistics" (Figure 2 Part C, data highlighted in green) for a specific confidence level (e.g. 95%). Maximum and minimum V_{50} values assuming a normal and a binomial error distribution can also be found. The difference between the STANAG 2920 and the normal distribution error as well as the description of other variables are presented graphically in Figure 4. It shows the fitted curve (e.g. Logit) in red along with its upper (light green) and lower (magenta) 95% percentile error curves based on normal distribution of the error. The confidence interval lengths, as defined in the STANAG 2920 are shown in blue, based on STANAG 2920 procedure and in yellow, based on the normal error curves.

https://doi.org/10.52202/078352-0035

As they are calculated in different ways using different distributions, the confidence interval lengths (CI), the maximum V_{50} values and the minimum V_{50} values are all different. Although not explicitly written, the V_{50} and σ values and their SE calculated using the NIJ 0101.06 procedure are highlighted in yellow in Figure 2 Part B.

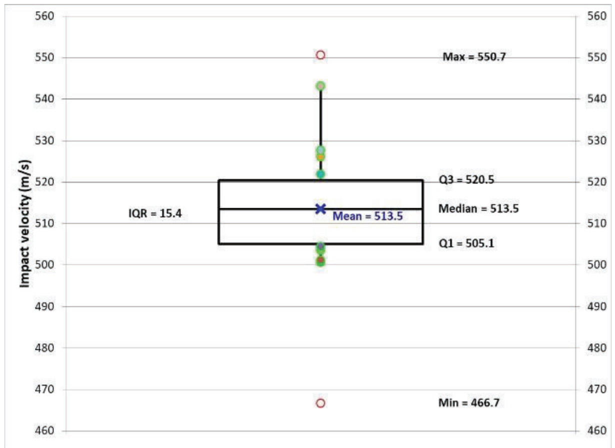

Figure 3 – Box plot page

4.4 Values and uncertainties related to any probability of perforation, including V_{proof}

Descriptions of the different values calculated by BLC are shown graphically in Figure 5. For a given probability of perforation (e.g. 10%), the associated impact velocity (V_{proof}), the minimum and maximum possible probability of perforation are displayed for either a normal error or a binomial error distribution. Similarly, the maximum expected V_{proof} value, based on the lower confidence limit curve is also displayed. These data can be found under the heading "V_{PROOF} Statistics", Figure 2 Part D. Notice that values are varying depending on the model used. Of particular interest in Figure 2 Part D is the error on the actual perforation probability. Although 5% probability is expected, the error of the model can result in probabilities of perforation as high as 38% for the Probit model with a 95% confidence level. This emphasises the use of V_{proof} values that corresponds to very low probability of perforation so that the maximum probability of perforation expected can be acceptable.

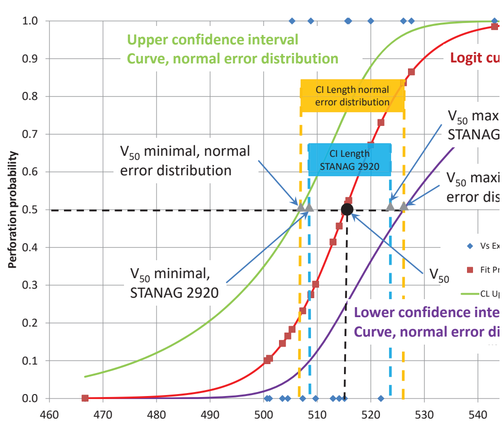

Figure 4 – Illustration of V_{50} statistics, extreme values and CI length

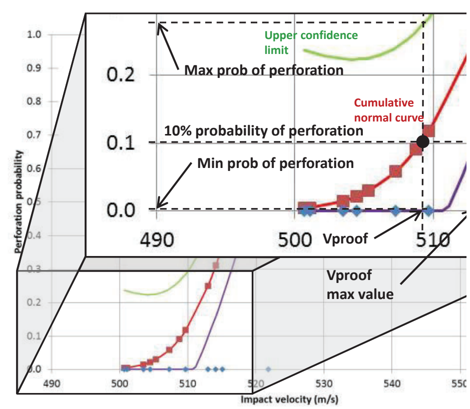

Figure 5 –Illustration of V_{proof} statistics with an example for 10% probability of perforation.

4.5 Identification of the best model or: Is the selected model valid?

https://doi.org/10.52202/078352-0035

This last assessment is a rather complex issue and necessitates the use of many statistics and tests described above. A flow chart of the process is presented in Figure 6.

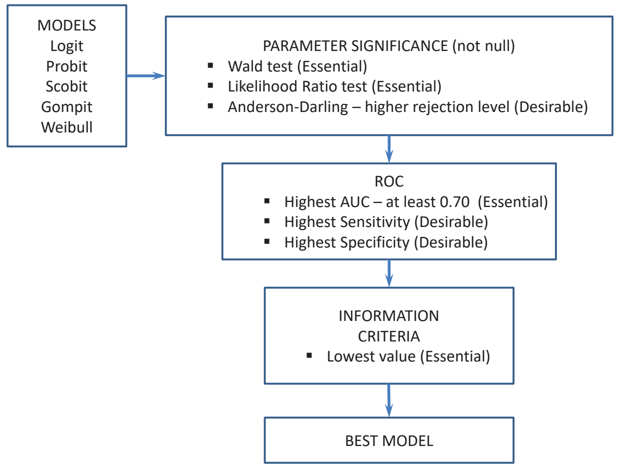

Figure 6 - Decision process to determine best model

To assess model validity, the parameter significance and the ROC criteria must be achieved. To assess which model best represents the experimental data, all 3 criteria (parameter significance, ROC and IC) must be achieved.

4.5.1 Parameter significance

The first thing is to verify if the calculated parameters ($\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$) are significant. If one of them is null, the model is worthless. To verify that point, the Wald test and Likelihood Ratio results can be analysed. Each of the 2 or 3 parameters have to be significant using the Wald Test. This enables to determine if each parameter is significant (not null). The Likelihood Ratio test compares the full model to the constant only ($\hat{\alpha}$ parameter only) model. The comparison enables to determine if the $\hat{\beta}$ parameter is significant (not null). The use of the AD test to assess if the experimental data sample is a good fit to the ML fitted link function can be done keeping in mind the conclusion relative to the acceptable level in section 3.4.

4.5.2 Goodness-of-fit

Goodness-of-fit of the model can be verified using the ROC. The ROC AUC has to be as close as possible to 1.0. A qualitative scale for the ROC AUC value is presented in [9]. Also, [9] recommends that the ROC AUC value be above 0.70, which corresponds to at least an acceptable level of discrimination. Because they represent the proportion of complete and partial perforation correctly identified by the model, sensitivity and specificity of the model has to be as close as possible to 1.

4.5.3 Information Criteria

Finally, the lowest the IC value, the best the model is. Of all the criteria presented in this section, this is the only one that can be used directly to compare between the different models. The variance on the IC value can be quite large. Therefore, when compared to each other's, the difference between the values of IC can be not significant. Reduction of the variance on the bias estimate can be achieved by increasing the number of experimental data point or by providing a more accurate evaluation of the bias and its variance by increasing the total number of Bootstrap samples used to assess variance. This can only be done at the cost of longer execution duration.

4.5.4 Example

In the present case, the 3 models should be considered valid (Figure 2 Part E). All the parameters are non-null as shown by the Wald and Likelihood ratio tests. The ROC AUC is 0.86, which means that each model presents excellent discrimination [9]. Sensitivity and specificity are also high, which means that the number of partial and complete perforation correctly identified by each model is high.

Assessment of the best model is more difficult. The IC values for the Probit is the lowest of the 3 models. On the other hand, the Scobit model have a large standard deviation value and as a result, the observed difference in the mean IC values for each model is not significant.

https://doi.org/10.52202/078352-0035

5. CONCLUSIONS

Capabilities of the BLC for the assessment of validity and goodness of fit of different models for V_{50} and V_{proof} assessment is explored. A variety of tests are discussed, accounting for the effects of small sample size.

6. REFERENCES

[1] AEP 2920 – Procedures for the Evaluation and Classification of Personal Armour, Bullet and Fragmentation Threats, Edition A, Version 1, 22 June 2015, NATO Standardization Office, Brussels.

[2] STANAG 2920 – Classification of Personal Armour (EDITION 3), 22 June 2015, NATO Standardization Office, Brussels.

[3] NIJ Standard-0101.06 – Ballistic Resistance of Body Armour, July 2008, U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

[4] Nagler, J., Scobit: An Alternative Estimator to Logit and Probit, American Journal of Political Science, Vol. 38, No. 1, February 1994, pp 230-255.

[5] Brathwaite, T. and Walker, J.L., Asymmetric, closed-form, finite-parameter models of multinomial choice, Journal of Choice Modelling (2017), http://dx.doi.org/10.1016/j.jocm.2018.01.002 .

[6] Mauchant, D., Rice, K.D., Riley, M.A., Leber, D., Samarov, D., and Forster, A.L., Analysis of Three Different Regression Models to Estimate the Ballistic Performance of New and Environmentally Conditioned Body Armor, U.S. Department of Commerce, National Institute of Standards and Technology, NISTIR 7760, February 2011.

[7] Nelder, J.A. and Mead, R.A., A simplex method for function minimization. Computer Journal, Vol. 7, 308–313, 1965.

[8] Williams, R., Maximum Likelihood Estimation, University of Notre Dame, http://www3.nd.edu/~rwilliam/, Last revised January 14, 2016, Accessed 13 July 2016.

[9] Hosmer, D.W. and Lemeshow, S., Applied Logistic Regression, Second Edition, Wiley Series on Probability and Statistics, 2000.

[10] Ku, H.H., Notes on the Use of Propagation of Error Formulas, Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation, Vol. 70C, No.4, October- December 1966.

[11] Lorenz, V., Yang, L., Grosse Perdekamp, M., Hertzog, D., Clegg, R., Error Analysis, https://courses.physics.illinois.edu/phys403/sp2016/lectures/ErrorAnalysis_Lorenz2016.pdf . PHYS403 Course lecture, University of Illinois at Urbana, College of Engineering, Spring 2016.

[12] Heo, J.H., Salas, J.D, Kim, K.D., Estimation of Confidence Intervals of Quantiles for the Weibull Distribution, Stochastic Environmental Research and Risk Assessment, Vol. 15, 2001, pp 284-309.

[13] Watkins, J., Topic 15 – Maximum Likelihood Estimation, University of Arizona, http://math.arizona.edu/~jwatkins/o-mle.pdf , November 1 and 3 2011. Accessed 13 July 2016.

[14] Das, U., Dhar, S.S. and Pradhan, V., Corrected likelihood-ratio tests in logistic regression using small-sample data, Communications in Statistics - Theory and Methods, 2017.

[15] Tanizaki, H., Power Comparison of Empirical Likelihood Ratio Tests - Small Sample Properties Through Monte Carlo Studies, Kobe University Economic Review, Vol. 50, 2004.

[16] Stephens, M.A., The Anderson-Darling Statistic, Prepared under Grant DAAG29-77-G-0O31 For the U.S. Army Research Office, Technical Report No. 39, October 31, 1979

[17] Razali, N.M. and Wah, Y.B., Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, Journal of Statistical Modeling and Analytics Vol.2 No.1, 21-33, 2011.

[18] Bispo, R., Marques. T.A. and Pestana, D., Statistical power of goodness-of-fit test based on empirical distribution function for Type-I right censored data, Journal of Statistical Computation and Simulation, 2011, pp 1-9.

[19] Cousineau, D. and Allen, T.A., Likelihood and its use in Parameter Estimation and Model Comparison, Mesure et évaluation en éducation, Vol 37, No. 3, pp 63–98, 2015.

https://doi.org/10.52202/078352-0035

[20] Anderson, D.R and Burnham, K.P., Understanding information criteria for selection among capture-recapture or ring recovery models, Bird Study, 46 (suppl.), 1999, pp S14-21.

[21] Konishi, S. and Kitagawa, G., Information Criteria and Statistical Modeling, Springer Series in Statistics, 2008, ISBN: 978-0-387-71886-6.

[22] Takeuchi, K., Distribution of informational statistics and a criterion of model fitting, Suri Kagaku (Mathematic Sciences), 153, 1976, pp.12-18 (in Japanese).

[23] Akaike, H., A new look at the statistical model identification. IEEE Transactions on Automatic Control, Vol. 19, No. 6, pp 716–723, 1974.

[24] Imori, S., Yanagihara, H., and Wakaki, H., General Formula of Bias-Corrected AIC in Generalized Linear Models, Scandinavian Journal in Statistics, Vol. 41,No. 2, pp 535–555, 2014.

https://doi.org/10.52202/078352-0035