

# Regression-based approaches for simulation meta-modelling in the presence of heterogeneity and correlation

L. P. D. M. Garces <sup>a</sup> , T. Bogomolov <sup>a</sup> , and B. A. Chiera <sup>a</sup> 

<sup>a</sup>UniSA STEM, University of South Australia, Adelaide, South Australia, 5000

Email: [len\\_patrick\\_dominic.garces@mymail.unisa.edu.au](mailto:len_patrick_dominic.garces@mymail.unisa.edu.au)

**Abstract:** We discuss several regression-based methods for simulation meta-modelling and illustrate these methods using combat simulator data. Since the use of common random numbers (CRNs) as a variance reduction technique induces correlations in the outputs generated by different simulation inputs, it is crucial to accommodate the possibility of heterogeneity, heteroskedasticity, and correlation when building meta-models. Furthermore, mainstream combat simulators produce a variety of output types, including continuous, binary, and count data. While extensive work has been done towards the development of simulation meta-modelling methods for continuous outputs, the meta-modelling of discrete, binary, and count data seems to be less understood.

To this end, we consider the use of estimated generalized least squares (EGLS), finite mixture generalized linear models (GLMs), and heteroskedastic binary regression, which specifically incorporate correlation, heteroskedasticity, and heterogeneity, for meta-modelling with continuous, binary, and count output data. EGLS extends the ordinary least squares (OLS) model by allowing the errors to have a non-diagonal covariance matrix. Finite mixture GLMs capture the possible heterogeneity in regression intercepts and slopes due to the possible existence of latent clusters in the simulation inputs. Heteroskedastic binary regression is a latent variable approach for binary data which jointly models the conditional mean and the scale parameter of the distribution of the latent error term.

An analysis of combat simulator data using the aforementioned methods shows that there is significant heterogeneity in the base mean levels and in the marginal effects of individual input variables for continuous and binary output data. Furthermore, likelihood ratio tests suggest an improved fit to the data when using heteroskedastic probit and logistic regression models over their homoskedastic counterparts. However, the analysis of count output data points to severe underdispersion in the data rather than heterogeneity in the sense of the finite mixture GLMs. This also suggests that approaches which jointly model the mean and dispersion may be viable alternatives.

**Keywords:** *Simulation meta-modelling, common random numbers, estimated generalized least squares, finite mixture generalized linear models, heteroskedastic binary regression*

## 1 INTRODUCTION

Simulation studies are commonly used to inform recommendations and support decision-making on competing outputs simulated under comparable conditions. Simulators used for this purpose, are executed for a total of  $m$  replications across  $n$  experiment design points, and generates a number of output metrics. The  $m$  replications in the simulation are designed to capture the stochastic variability in simulation conditions. As  $m$  and  $n$  increase, simulation analysis can become enormously time-consuming. Depending on the configuration of design points, the run time of a single replication of mainstream combat simulators can span from hours to days. The use of meta-models to obtain a functional approximation of the relationship between the black-box simulation input (design points) and output performance responses (metrics) addresses this issue. Common random numbers (CRNs) are typically used as a variance-reduction technique for simulation experiments. In this approach the same pseudo-random number stream is used for each of the design points to subject all scenarios to the same statistical environment, allowing for a fair comparison between distinct design points. It is well-known however that the use of CRNs in simulation experiments induces correlation in the outputs generated by distinct design points (see e.g. Kleijnen, 1992; Gill et al., 2018), thereby complicating analyses using ordinary least squares (OLS) regression or generalized linear models (GLMs).

A further complication lies in the nature of the output which may yield metrics which are categorical, continuous, and discrete. Meta-modelling approaches for continuous output metrics have been extensively covered in the literature (see e.g. Chen et al., 2009, for a review of these methods). However, there appears to be very little attention paid to simulation meta-modelling with binary, discrete, or count output metrics. Meckesheimer et al. (2001) tackles the problem of meta-modelling with piecewise-continuous responses, where breaks are dictated by some other discrete output, but their approach does not accommodate the meta-modelling of strictly binary, discrete, or count outputs. Furthermore, Gill et al. (2018) assumed that metrics followed a continuous distribution to enable the use of OLS regression, but such an approach for modelling binary, discrete, or count outputs is problematic.

This paper contributes to the simulation meta-modelling literature by discussing a framework for a regression-based meta-modelling of simulations with continuous, binary, and count outputs. The methods discussed in this paper also account for possible heterogeneity and correlation induced by the use of CRNs. Specifically, we illustrate the application of estimated generalized least squares (EGLS) (Kleijnen, 1992), finite mixture GLMs (Wedel and DeSarbo, 1995), and heteroskedastic binary regression (Alvarez and Brehm, 1995) and compare these results to those obtained from the usual OLS or GLM models. By taking into account any possible heteroskedasticity, heterogeneity, or correlation in the data, we obtain more statistically reliable estimates and inferences. These can then be used for other purposes such as sensitivity analysis, design point comparison, and ranking of alternatives, especially in defence-related situations. We prefer regression-based approaches since these structures are more interpretable compared to other competing methods.

The rest of the paper is organized as follows. Section 2 briefly discusses the EGLS, finite mixture GLM, and heteroskedastic binary regression methods. We then apply these methods in Section 3 using actual combat simulator data. Section 4 synthesizes the findings in the paper.

## 2 DISCUSSION OF REGRESSION-BASED APPROACHES

### 2.1 Estimated Generalized Least Squares

In this section, we review the EGLS method proposed by Kleijnen (1992). We adopt the following notation in this section. For each design point  $i$ ,  $i = 1, \dots, n$ , we let  $\mathbf{x}_i = (1, x_{i,1}, \dots, x_{i,L})^\top$  denote the regressors (possibly including interaction or higher-order terms) and  $\{w_{i,r}\}_{r=1}^m$  the  $m$  realizations of the continuous output variable  $w_i$ . Let  $\bar{\mathbf{w}} = (\bar{w}_1, \dots, \bar{w}_n)^\top$ , where  $\bar{w}_i = \frac{1}{m} \sum_{r=1}^m w_{i,r}$ , be the vector of simulation output averages. Let  $\mathbf{X}$  be the  $n \times (L + 1)$  matrix whose  $i$ th row is  $\mathbf{x}_i^\top$ . We denote by  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_L)^\top$  the unknown regression coefficients and by  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$  the vector of error terms.

The generalized least squares (GLS) framework extends the OLS model by accommodating possible heteroskedasticity and correlated errors. The GLS model assumes the linear input-output relationship  $w_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i$ , and assumes that  $\text{Var}[\boldsymbol{\epsilon}|\mathbf{X}] = \Sigma$ , where  $\Sigma = [\sigma_{i,j}]$  is a usually unknown  $n \times n$  positive semi-definite matrix with possibly unequal main diagonal entries and nonzero off-diagonal entries. The GLS estimator of  $\boldsymbol{\beta}$  is given by  $\hat{\boldsymbol{\beta}}^{GLS} = (\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Sigma^{-1} \bar{\mathbf{w}}$ , which can be calculated if  $\Sigma$  is known. Replicated measurements of the output allow us to estimate  $\Sigma$  via the sample covariance matrix  $\hat{\Sigma} = [\hat{\sigma}_{i,j}]$ , where  $\hat{\sigma}_{i,j} = \frac{1}{m-1} \sum_{r=1}^m (w_{i,r} - \bar{w}_i)(w_{j,r} - \bar{w}_j)$ , for  $i, j = 1, \dots, n$ . Using the estimated covariance matrix in the

GLS estimator yields the EGLS estimator of  $\beta$ , which is given by

$$\hat{\beta}^{EGLS} = (\mathbf{X}^\top \tilde{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \tilde{\Sigma}^{-1} \tilde{\mathbf{w}} \quad (1)$$

where  $\tilde{\Sigma} = \hat{\Sigma}/m$  and  $\hat{\Sigma} = [\hat{\sigma}_{i,j}]$ . This process requires  $m > n$  since  $\hat{\Sigma}$  is singular otherwise (Kleijnen, 2015).

Confidence intervals for the regression coefficients based on the EGLS estimator can be constructed via *jackknifing*. Jackknifing involves, for each  $r = 1, \dots, m$ , calculating the EGLS estimator  $\hat{\beta}^{J:(-r)}$  after removing the  $r$ th replication. For each regression coefficient  $\beta_\ell$ ,  $\ell = 0, 1, \dots, L$ , we calculate the  $m$  pseudo-values  $J_{\ell;r} = m\hat{\beta}_\ell^{EGLS} - (m-1)\hat{\beta}_\ell^{J:(-r)}$  (where  $\hat{\beta}_\ell^{EGLS}$  is the  $\ell$ th element of the EGLS estimator given in (1)), for  $r = 1, \dots, m$  and calculate the average pseudo-value  $\bar{J}_\ell = \frac{1}{m} \sum_{r=1}^m J_{\ell;r}$ . Assuming that the pseudo-values are i.i.d. with a normal distribution, the endpoints of the  $100(1 - \alpha)\%$  jackknifed confidence interval for  $\beta_\ell$  are

$$\bar{J}_\ell \pm t_{1-\frac{\alpha}{2}, m-1} \times S_\ell^J, \quad \text{where } S_\ell^J = \sqrt{\frac{\sum_{r=1}^m (J_{\ell;r} - \bar{J}_\ell)^2}{m(m-1)}},$$

and  $t_{1-\frac{\alpha}{2}, m-1}$  is the critical  $t$ -value with left-tail probability  $1 - \frac{\alpha}{2}$  and  $m - 1$  degrees of freedom. Since EGLS estimation is performed after removing one replication at a time, jackknifing requires  $m - 1 > n$  to ensure that all the estimated covariance matrices are nonsingular. Alternatively, one can construct confidence intervals using the asymptotic covariance matrix of the EGLS estimator, but this is problematic for small sample sizes. Bootstrapping procedures can also be used to construct confidence intervals (see e.g. Kleijnen, 2015, Section 3.5), but we do not consider this method in this analysis.

## 2.2 Finite Mixture Generalized Linear Models

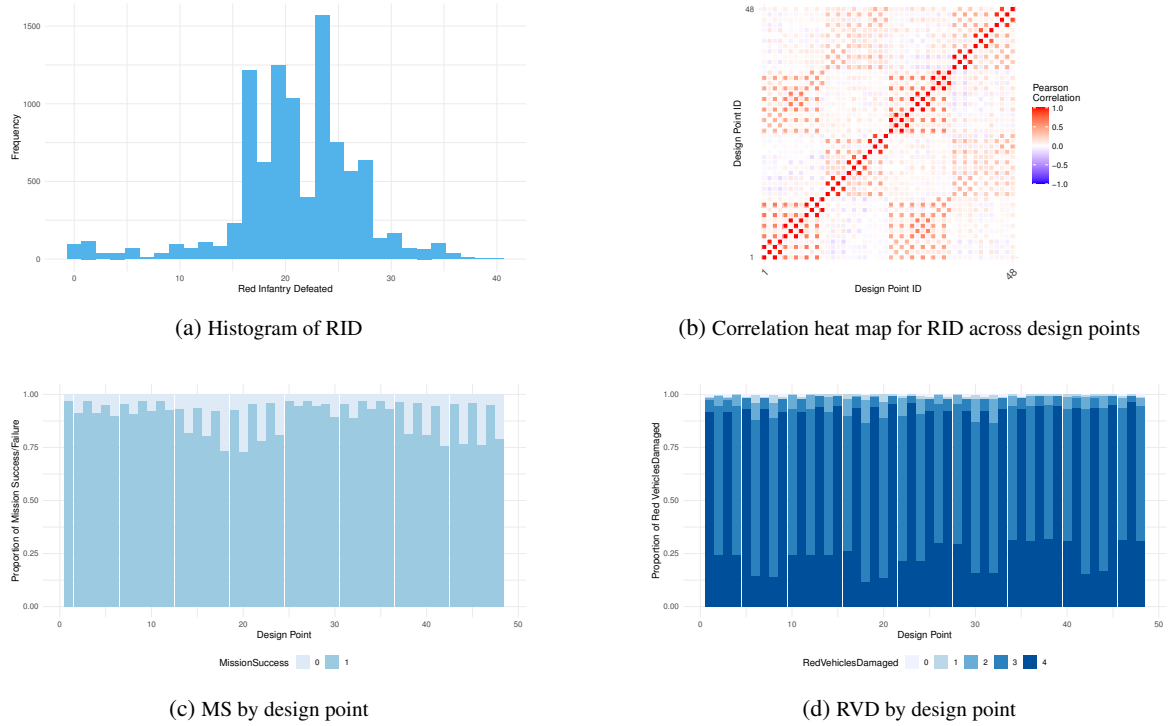
For non-continuous output metrics, the notion of ‘‘correlation’’ is not as well-defined as it is the case for continuous outputs. As such, even with access to replicated output measurements for each design point, it becomes difficult to assess how much correlation is induced by CRNs. Thus, in the case of non-continuous outputs, we instead consider the possibility of *heterogeneity* induced by CRNs. Here, we define heterogeneity as the situation where there may exist smaller unknown subsets or *latent classes* of the unique design points where the base effect (i.e. regression intercept) or the main effect of input variables (i.e. regression coefficients/slopes) may vary across the latent classes. This heterogeneity can be modelled using a finite mixture of GLMs as proposed by Wedel and DeSarbo (1995). This approach accommodates any output type, e.g. continuous, binary, and count output, that can be modelled using a member of the exponential family.

Finite mixture GLMs are specified as follows. Taking  $S$  to be the number of latent classes, for each unique design point  $i$  we define the vector  $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,S})$ , where  $u_{i,s} = 1$  if design point  $i$  belongs to class  $s$ ,  $s = 1, \dots, S$ . For each  $i$ , the  $u_{i,s}$ ’s are assumed to be i.i.d. multinomial with probabilities  $\alpha_s$ , i.e.  $\mathbf{u}_i$  has joint probability mass function  $p(\mathbf{u}_i; \alpha) = \prod_{s=1}^S \alpha_s^{u_{i,s}}$ , where  $\alpha = (\alpha_1, \dots, \alpha_S)$ . We note that the  $\mathbf{u}_i$ ’s are unobserved and the  $\alpha_s$ ’s satisfy  $\sum_{s=1}^S \alpha_s = 1$  and  $\alpha_s \geq 0$  for all  $s$ . Conditional on design point  $i$  belonging to latent class  $s$ , the output  $w_i$  has conditional probability density/mass function  $f_i^{(s)}(w; \beta_s)$ , which belongs to the exponential family of distributions. Here,  $\beta_s = (\beta_{0,s}, \beta_{1,s}, \dots, \beta_{L,s})^\top$  are the regression parameters for latent class  $s$ . The conditional mean  $\mu_i^{(s)} = E[w_i | \mathbf{x}_i, u_{i,s}]$  is related to the linear predictor via  $g(\mu_i^{(s)}) = \mathbf{x}_i^\top \beta_s$ , where  $g$  is an appropriate link function. Let  $\beta = (\beta_1^\top, \dots, \beta_S^\top)$  be the collection of all regression parameters across all latent classes. The conditional probability density/mass function of  $w_i$  given  $\mathbf{u}_i$  is therefore  $f_i(w | \mathbf{u}_i; \beta) = \prod_{s=1}^S [f_i^{(s)}(w; \beta_s)]^{u_{i,s}}$ . The joint probability density function of  $w_i$  and  $\mathbf{u}_i$  is  $p(w_i, \mathbf{u}_i; \alpha, \beta) = f_i(w_i | \mathbf{u}_i; \beta) p(\mathbf{u}_i; \alpha)$ , so the *complete* log-likelihood function is given by

$$L_c(\alpha, \beta) = \ln \prod_{i=1}^n p(w_i, \mathbf{u}_i; \alpha, \beta) = \sum_{i=1}^n \sum_{s=1}^S u_{i,s} \ln f_i^{(s)}(w_i; \beta_s) + \sum_{i=1}^n \sum_{s=1}^S u_{i,s} \ln \alpha_s. \quad (2)$$

Since the likelihood (2) involves unobserved data  $\mathbf{u}_i$ , the maximum likelihood estimation of the parameters is carried out using the EM algorithm (Dempster *et al.*, 1977; Wedel and DeSarbo, 1995).

The number of classes  $S$  must also be estimated from the data. In this analysis, we estimate a finite mixture model for each value of  $S$  in a pre-specified set of values and choose the  $S$  which maximizes the log-likelihood or minimizes the AIC or the BIC. The estimation of finite mixture models in this analysis is implemented using the **flexmix** package in R (Leisch, 2004).



**Figure 1.** Visual summaries of the output metrics: number of red infantry defeated (RID), mission success (MS), and number of red vehicles damaged (RVD)

### 2.3 Heteroskedastic Binary Regression via Latent Variables

We first consider the heteroskedastic probit model. Suppose we observe a binary outcome  $w_i$ , which we assume is given by  $w_i = \mathbf{1}(y_i^* > 0)$ , where  $\mathbf{1}(\cdot)$  is the indicator function and  $y_i^*$  is a latent variable of the form  $y_i^* = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i$ . Let  $p_i = \Pr(w_i = 1 | \mathbf{x}_i) = \mathbb{E}[w_i | \mathbf{x}_i]$ . Instead of assuming that the  $\epsilon_i$ 's are i.i.d.  $N(0, 1)$ , we assume  $\epsilon_i \sim N(0, \sigma_i^2)$ , reflecting heteroskedasticity in the latent error term. The heteroskedastic probit model is specified as

$$\Phi^{-1}(p_i) = \frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma_i}, \quad h(\sigma_i) = h(1) + \mathbf{z}_i^\top \boldsymbol{\gamma}, \quad (3)$$

where  $h(\sigma_i) = h(1) + \mathbf{z}_i^\top \boldsymbol{\gamma}$  is the scale model, with  $h$  a link function (usually the log, square root, or identity function),  $\mathbf{z}_i^\top$  a vector of regressors (not necessarily equal to  $\mathbf{x}_i$ ), and  $\boldsymbol{\gamma}$  a vector of unknown coefficients in the scale model. Aside from the probit link function  $g(p) = \Phi^{-1}(p)$ , one can also use the usual logit link function  $g(p) = \ln(p/(1-p))$  in (3) (see Koenker and Yoon (2009) for other appropriate link functions). The **glm**x package in R (Zeileis et al., 2015) provides a suite of functions for fitting heteroskedastic binary regression models.

## 3 ILLUSTRATION WITH COMBAT SIMULATOR DATA

In this section, we illustrate the implementation of the aforementioned statistical methods to the meta-modelling of a combat simulator. In all regression runs, we only consider the main effects, i.e. an intercept term and linear terms. Nonetheless, the analysis can easily be extended to include higher-order terms such as quadratic or interaction terms. We assume a 5% level of significance for all hypothesis tests and confidence intervals. Variables marked with \* are statistically significant at  $\alpha = 5\%$ . All statistical analyses were performed using established utilities or packages in R, so the computational time for all runs is negligible.

In the supplied data set, there are three output variables, namely *mission success* (binary), *number of red infantry defeated* (count), and *number of red vehicles damaged* (count). The simulation inputs consist of four discrete variables, namely *Option* (“A”, “B”, “C”, or “D”), *F1* (“Direct” or “Indirect”), *F2* (“25” or “75”), and *F3* (“Low”, “Medium”, or “High”). There are  $n = 48$  distinct input combinations, each of which corresponds to  $m = 200$  replicated measurements of each of the output variables. The combat simulator used to generate

**Table 1.** OLS, EGLS, and finite mixture Gaussian regression estimates for number of red infantry defeated.

	OLS		EGLS				FM Gaussian Regression					
	(HC se)		Est.	95% Jackknifed CI		Class 1	Class 2	Class 3				
Intercept	23.340	*	23.454	*	23.039	24.489	22.981	*	22.405	*	23.234	*
Option B	-1.029	*	-1.255	*	-2.338	-0.789	-2.607	*	0.385	*	-0.952	*
Option C	-0.105		-0.035		-0.795	0.278	0.177	*	-0.069		1.140	*
Option D	-1.119	*	-1.193	*	-2.263	-0.744	-1.833	*	-0.194	*	-1.066	*
F1 Indirect	-2.839	*	-2.933	*	-3.808	-2.426	-1.779	*	-1.891	*	-2.789	*
F2 75	0.009		0.007		-0.001	0.070	-0.064	*	0.007		-0.228	*
F3 Low	-0.816	*	-0.647	*	-1.044	-0.166	-1.408	*	-0.191	*	-1.158	*
F3 Medium	-0.280	*	-0.309	*	-0.820	-0.071	-0.349	*	0.075	*	-0.738	*
Log-Lik	-30433.70						54.68					
AIC	60885.41						-51.37					
BIC	60949.94						2.89					

the data set takes between 25 to 345 minutes (with an average run time of 185 minutes) to produce the output metrics for a single repetition for each unique input combination. For illustrative purposes we consider the *number of red infantry defeated* a continuous output rather than a count output since, as shown in Figure 1a, it has a relatively high average value, good dispersion, and an approximately symmetric distribution.

### 3.1 Number of Red Infantry Defeated

Table 1 shows the OLS, EGLS, and finite mixture Gaussian regression estimates, as well as the 95% jackknifed confidence interval corresponding to the EGLS estimates. Inferences for the OLS model are performed using heteroskedasticity-consistent standard errors since the Breusch-Pagan test shows that there is heteroskedasticity in the data ( $p$ -value  $< 2.2e^{-16}$ ). Figure 1b shows that RID measurements are highly correlated across “adjacent” design points. The estimated covariance matrix also shows strong heterogeneity in the output variance across design points. These imply that EGLS is a more appropriate method for constructing a linear regression model. Inferences on the EGLS coefficients are based on the jackknifed confidence intervals and conclusions on direction of effect and statistical significance similar to those from the OLS model are obtained for the EGLS model. However, the OLS model returns an adjusted  $R^2$  of 6.654%, indicating a severe lack of fit to the data. Rao’s lack-of-fit  $F$ -test (Kleijnen, 1992) applied to the EGLS model also points to a lack of fit to the data (test statistic: 1.69,  $p$ -value: 0.0041).

We fit a finite mixture of Gaussian models (OLS) to the RID data for  $S = 1, \dots, 10$  latent classes and found that  $S = 3$  yields the highest likelihood and lowest AIC and BIC. Using a finite mixture model results in a drastic improvement in the log-likelihood, AIC, and BIC compared to the OLS model. We note further that there are input variables whose effect on RID changes across clusters. For example, the magnitude *and* sign of Option B’s effect changes across clusters.

### 3.2 Mission Success

Figure 1c shows that simulated missions are overwhelmingly successful, which may produce bias towards successful missions if the entire data set is used to estimate the binary regression models. Therefore, all models in this subsection are estimated using a balanced data set consisting of 600 successful and 600 failed missions. The sampling strategy ensures that all design points and replications are adequately represented. Accuracy (AR) and balanced accuracy ratios (Bal. AR) are calculated using the full data set to assess the models’ classification performance. In Table 2, we show the estimates of the homoskedastic and heteroskedastic probit and logit binary regression and the finite mixture logistic regression. For the heteroskedastic binary regression, we used the same regressors in the scale model as in the mean model, i.e. we set  $\mathbf{x}_i = \mathbf{z}_i$  in (3). Only the mean model of the heteroskedastic probit and logit models are presented, since none of the regressors are significant in the scale model.

We note the identical classification performance of the homoskedastic probit and logistic regression, as well as identical conclusions about the sign and significance of the main effects. When including a model for the scale parameter, some variables lose statistical significance in the mean model. Furthermore, the likelihood ratio test (LRT) comparing the homoskedastic and heteroskedastic models indicate a significant improvement in the fit when the heteroskedastic model is used. More importantly, there seems to be mixed results regarding the classification performance of the heteroskedastic models. Both heteroskedastic models yield a balanced accuracy of 50%, while the reported accuracy ratio is 90% since the heteroskedastic models classified all observations in the full data set as successful missions. Thus, while there may be evidence for heteroskedasticity in the

**Table 2.** Homoskedastic and heteroskedastic binary regression (with probit and logit link functions) and finite mixture logistic regression results for mission success.

	Homoskedastic		Heteroskedastic		FM Logistic Regression	
	Probit	Logit	Probit	Logit	Class 1	Class 2
Intercept	1.2041 *	2.0124 *	0.8818 *	1.4441 *	0.8785 *	1.0220 *
Option B	-0.8926 *	-1.4877 *	-0.4215	-0.7117	-0.2800 *	-0.4215 *
Option C	-0.1523	-0.2569	-0.0361	-0.0623	0.0320	-0.3944 *
Option D	-0.7843 *	-1.3804 *	-0.2456	-0.4194	-0.2879 *	-0.2895
F1 Indirect	-0.8402 *	-1.3954 *	-0.7586 *	-1.2322 *	-0.2872 *	-0.3758 *
F2 75	-0.0772	-0.1370	-0.0441	-0.0787	-0.0199	-0.0321
F3 Low	-0.2634 *	-0.4357 *	-0.0961	-0.1632	-0.0934 *	0.0280
F3 Medium	0.0352	0.0616	0.0038	0.0054	0.0010	0.0877
Log-Lik	-713.01	-712.38	-703.60	-703.60	-744.37	
AIC	1442.00	1440.80	1437.30	1437.27	1526.75	
BIC	1482.73	1481.47	1513.65	1513.62	1623.46	
LR Test p-val.			0.0091	0.0145		
AR	0.7114	0.7114	0.9024	0.9024		
Bal. AR	0.6716	0.6716	0.5000	0.5000		

mission success, using these models may result in inaccurate predictions about mission outcomes. The finite mixture model provides some evidence of heterogeneity in the main effects (see e.g. Option C and Option D), but yields a worse log-likelihood and AIC/BIC compared to the homoskedastic and heteroskedastic logistic regression models.

### 3.3 Number of Red Vehicles Damaged

The distribution of RVD across design points as shown in Figure 1d points to a high degree of *underdispersion* (the variance of RVD is significantly less than the mean) in the data and may suggest that the Poisson regression may not be appropriate. As such, we also consider other count data models that accommodate under- or overdispersion, such as the quasi-Poisson, the Conway-Maxwell-Poisson (CMP), the negative binomial, and the Poisson-lognormal mixture models. For further details on these alternative count data regression methods, see e.g. Cameron and Trivedi (2013, Chapter 4). Table 3 shows the estimates of the regression coefficients and the dispersion parameter (for the Poisson, quasi-Poisson, and the CMP models). Attempts to fit a finite mixture Poisson model on the RVD data yields  $S = 1$  as the optimal number of clusters, therefore leading back to the ordinary Poisson regression model. In this case, there does not seem to be any evidence of heterogeneity in terms of the effects of the input variables on the mean Poisson rate of RVD.

A formal dispersion test indicates that the data is substantially underdispersed and is corroborated by the results of the quasi-Poisson regression. Regression coefficient estimates for the quasi-Poisson model are similar to those obtained in the ordinary Poisson regression, but the standard errors are substantially smaller, leading to more variables becoming statistically significant. These inferences are also identical to those derived from the CMP model. We observe that the signs of the regression coefficients are the same for all regressors across the three methods. The results from the negative binomial regression, specifically the regression coefficient estimates, individual tests of significance, and the log-likelihood, are identical to those from the usual Poisson regression. This is expected since the negative binomial model, when fit to underdispersed data, converges to the Poisson model. Similar conclusions about the main effects and individual tests of significance can also be made regarding the Poisson-lognormal model.

## 4 CONCLUSIONS

This paper discusses various regression-based approaches for the meta-modelling of simulation experiments with continuous, binary, and count output metrics. The methods discussed are able to handle heteroskedasticity, correlation, and heterogeneity which arise from the usage of CRNs in the simulator. Our analysis using finite mixture GLMs has shown that there is significant heterogeneity in the base and main effects to the mean of the continuous and binary outputs. The results arising from the finite mixture GLMs, however, must be appraised via a qualitative assessment of the design points that are clustered together in latent classes to determine why such a clustering was derived from the data. We also considered methods which specifically tackle heteroskedasticity in binary data (via a latent variable approach) and dispersion in count data and showed how accommodating these phenomena may change the conclusions when testing input variables for statistical significance.

**Table 3.** Results of the Poisson (P), quasi-Poisson (QP), Conway-Maxwell Poisson (CMP), negative binomial (NB), and Poisson-lognormal mixture (PLN) regression for number of red vehicles damaged.

	P		QP		CMP		NB		PLN	
Intercept	1.3621	*	1.3621	*	15.8125	*	1.3621	*	1.3723	*
Option B	0.0023		0.0023		0.0219		0.0023		-0.0053	
Option C	0.0084		0.0084		0.0813		0.0084		0.0081	
Option D	0.0184		0.0184	*	0.1765	*	0.0184		0.0111	
F1 Indirect	-0.2186	*	-0.2186	*	-2.0633	*	-0.2186	*	-0.2228	*
F2 75	0.0007		0.0007		0.0064		0.0007		0.0003	
F3 Low	-0.0288	*	-0.0288	*	-0.2760	*	-0.0288	*	-0.0274	*
F3 Medium	-0.0018		-0.0018		-0.0165		-0.0018		0.0005	
Log-lik	-15488.30				-8752.54		-15488.33		-15299.16	
AIC	30993.00				17523.08		30995.00			
BIC	31049.97				17587.60		31059.18		-15340.40	
Dispersion	0.0874		0.0875		0.0932					

As it is likely that there is some degree of under- or overdispersion in the output metrics, especially for count data, simulation meta-modelling via a joint modelling of the mean and dispersion (see e.g. Smyth, 1989) may be a useful alternative. In addition, for count data with a known upper limit (as is the case for most defence-related applications), approaches related to truncated or censored count data regression may be considered (see e.g. Cameron and Trivedi, 2013, Sections 4.3-4.4). Furthermore, additional work must be done to properly characterize the extent to which CRNs induce “correlation” in binary and discrete output metrics. A well-defined notion of “correlation” would then inform the selection of the appropriate statistical methodology for simulation meta-modelling.

#### ACKNOWLEDGMENTS

The Commonwealth of Australia (represented by the Defence Science and Technology Group) supported this research through a Defence Science Partnership agreement. The authors also thank Dr. Andrew Gill for fruitful discussions about the data set and the statistical methodology employed in this paper.

#### REFERENCES

- Alvarez, R.M., Brehm, J., 1995. American ambivalence towards abortion policy: development of a heteroskedastic probit model of competing values. *American Journal of Political Science* 39, 1055–1082.
- Cameron, A.C., Trivedi, P.K., 2013. *Regression Analysis of Count Data*. 2nd ed., Cambridge University Press, New York.
- Chen, V.C.P., Tsui, K.L., Barton, R.R., Meckesheimer, M., 2009. A review on design, modeling and applications of computer experiments. *IIE Transactions* 38, 273–291.
- Dempster, A.P., Laird, N.M., Rubin, R.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B* 39, 1–38.
- Gill, A., Grieger, D., Wong, M., Chau, W., 2018. Combat simulation analytics: Regression analysis, multiple comparisons and ranking sensitivity, in: Rabe, M., Juan, A.A., Mustafee, N., Skoogh, A., Jain, S., Johansson, B. (Eds.), *Proceedings of the 2018 Winter Simulation Conference*, IEEE Press. p. 3789–3800.
- Kleijnen, J.P.C., 1992. Regression metamodelling for simulation with common random numbers: comparison of validation tests and confidence intervals. *Management Science* 38, 1164–1185.
- Kleijnen, J.P.C., 2015. *Design and Analysis of Simulation Experiments*. 2nd ed., Springer International Publishing, Switzerland.
- Koenker, R., Yoon, J., 2009. Parametric links for binary choice models: a Fisherian-Bayesian colloquy. *Journal of Econometrics* 152, 120–130.
- Leisch, F., 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software* 11, 1–18.
- Meckesheimer, M., Barton, R.R., Simpson, T.W., Limayem, F., Yannou, B., 2001. Metamodeling of combined discrete/continuous responses. *AIAA Journal* 39, 1950–1959.
- Smyth, G.K., 1989. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society, Series B (Methodological)* 51, 47–60.
- Wedel, M., DeSarbo, W.S., 1995. A mixture likelihood approach for generalized linear models. *Journal of Classification* 12, 21–55.
- Zeileis, A., Koenker, R., Doebler, P., 2015. *glm: Generalized Linear Models Extended*. R package version 0.1-1.