Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

EXPLOITING FUNCTIONAL DATA FOR COMBAT SIMULATION SENSITIVITY ANALYSES

Lisa Blumson[1], Charlie Peter[1], and Andrew Gill[1]

[1]Defence Science and Technology Group, Edinburgh, SA, AUSTRALIA

ABSTRACT

Computationally expensive combat simulations are often used to inform military decision-making and sensitivity analyses enable the quantification of the effect of military capabilities or tactics on combat mission effectiveness. The sensitivity analysis is performed using a meta-model approximating the simulation's input-output relationship and the output data that most combat meta-models are fitted to correspond to end-of-run mission effectiveness measures. However during execution, a simulation records a large array of temporal data. This paper seeks to examine whether functional combat meta-models fitted to this temporal data, and the subsequent sensitivity analysis, could provide a richer characterization of the effect of military capabilities or tactics. An approach from Functional Data Analysis will be used to illustrate the potential benefits on a case study involving a closed-loop, stochastic land combat simulation.

1 INTRODUCTION

Combat simulations are often used to inform military decision-making, as real-world data collection is expensive or dangerous, and form part of a multi-method approach to quantify the operational effectiveness of alternative military capabilities or tactics. These simulations complement other methods such as war-gaming, field trials and historical analysis while allowing for the exploration of a larger parameter space and the ability for a scenario to be replicated many times in order to conduct statistical comparisons. Data from events related to movement, acquisition, fires, communications and resource consumption can be captured and analyzed to investigate the cause-and-effect relationships in military environments. AFSIM (Zeh et al. 2014) operates at the engagement and mission level and models detailed weapon kinematics, along with various sensor, communication and battle management systems. MANA is a combat simulation which employs an agent-based approach and lower-fidelity modeling in order to allow emergent behavior to appear from the stochastic interactions between entities in a tactical setting (Lauren and Stephen 2002). COMBATXXI is a combat simulation used extensively in the authors' area for weapon systems and tactics evaluation for brigade and below combined arms conflicts, and models various battlefield functions such as air defense, amphibious warfare, aviation, close combat, combat service support, counter-mobility, and logistics (Balogh and Harless 2003). Finally, JICM (Ong and Ling 2002) and more recently, STORM (Seymour 2014) are combat simulations developed to analyze entire military campaigns and thus model across the land, air and maritime domains.

Computationally expensive combat simulations such as those above, are often replaced by simpler, approximate models of their input-output relationship for the purpose of prediction, optimization, or sensitivity analysis (Santner et al. 2003). This model of a model (meta-model) is fit from simulation data generated according to a designed experiment. For combat simulation sensitivity analyses, an appropriate meta-model can characterize varying a simulation factor through its impact (main effect) and synergies (interactions with other factors) (Sanchez et al. 2012) and thus enables the quantification of the effect of military capabilities or tactics on combat mission effectiveness (Gill et al. 2018). Typically, the output data that combat meta-models are fitted to correspond to end-of-run mission effectiveness measures (e.g., time to complete mission; did Blue win; number of surviving forces) and standard model fitting routines are well-known, such as linear, logistic and Poisson regression (Dunn and Smyth 2018). Thus, typically, the effect of the military factors is characterized and quantified relative only to the eventual combat outcome. However, there is an abundance of temporal data that combat simulations produce describing the state of combat at each of a discrete number of time-points (either equi-spaced if the simulation is time-driven, or not if the simulation is event-driven). The motivation behind this paper is to examine whether a combat meta-model fitted to this functional output data, and the associated combat simulation sensitivity analysis, might provide a more complete characterization of the effect of military factors, and thus improved analytical support to military decision-making. Specifically, we will investigate the use of functional linear meta-modeling to derive functional combat simulation meta-models.

Functional linear meta-modeling sits within Functional Data Analysis (FDA), an expanding field of statistics aimed at analyzing data that exists over some continuum (Ramsay and Silverman 2005). According to Ramsay and Silverman (2005), FDA methods treat observed data as estimates of a single functional entity, as opposed to merely a sequence of individual observations. FDA techniques have been applied across a number of diverse and distinct research areas, including the analysis of functional response data from simulations of designed experiments (Scherer 2021). This is an information-rich approach to simulation analysis, capitalizing on the efficiency of designed experiments while reducing information loss by analyzing the factor effects across the entire response curve as opposed to only statistical summaries of this longitudinal data. Scherer (2021) demonstrates the additional breadth of insight and knowledge that can be gained through a functional analysis of factor effects over time.

In this paper, we seek to bring to the attention of the simulation analytics community the concepts and methods of functional linear meta-modeling. To the authors' knowledge, this has not been applied to combat simulations nor published often at the Winter Simulation Conference. The potential benefits over typical combat sensitivity analyses are illustrated using a closed-loop, stochastic land combat simulation, focusing on combat power over time functional data.

2 BACKGROUND

2.1 Combat Simulation Sensitivity Analysis

The most common meta-model for the sensitivity analysis of stochastic combat simulations is a fully second-order response surface model, given by

$$y_{i,r} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} + \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} \beta_{j,k} x_{i,j} x_{i,k} + \varepsilon_{i,r}, \quad i = 1, \ldots, n \tag{1}$$

where $y_{i,r}$ is the end-of-run response when the simulation is replicated for the r^{th} time with the p factors set at the i^{th} combination of levels $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})$. Typically the errors $\varepsilon_{i,r}$ are assumed to be independent and identically distributed ($\mathrm{E}(\varepsilon_{i,r}) = 0$, $\mathrm{Var}(\varepsilon_{i,r}) = \sigma^2$) in which case Ordinary Least Squares (OLS) can be used to estimate the meta-model parameters via

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\bar{\mathbf{y}}, \quad \mathrm{Var}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{R}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \tag{2}$$

where $\hat{\boldsymbol{\beta}}$ is the stacked vector of estimated factor effects ($\hat{\beta}_j$ and $\hat{\beta}_{j,k}$), $\mathbf{X}$ is the model matrix formed by extending the design by appending element-wise products of the design's columns $\mathbf{x}_j = (x_{1,j}, \ldots, x_{n,j})^{\mathrm{T}}$, T denotes the transpose, and $\bar{\mathbf{y}}$ is the vector of responses ($\bar{y}_i$) averaged over all replications (see Kleijnen (2015) and Gill (2021) for details, including variations for when the errors cannot be assumed independent or identically distributed). Standard hypothesis testing procedures are then employed to determine the statistical significance of each of the estimated meta-model terms. For meta-models (1) and (2), the design underpinning $\mathbf{X}$ often comes from two-level fractional factorial designs of resolution at least V (Montgomery 2012), as these are both effective and efficient under typical optimality criteria (Fedorov 1972).

2.2 Functional Linear Meta-Modeling

Functional linear meta-models have a similar structural form to the general linear model in (1) with the distinction that the response, factors, effect estimates and errors may exist as functions over some continuum (Ramsay and Silverman 2005). There are three main classes of functional linear models depending on where in the model the functionality occurs; scalar-on-function, function-on-scalar and function-on-function. As simulation meta-models are typically derived from designed experiments with a scalar design matrix, focus here is kept on the function-on-scalar regression (FoSR). In a typical FoSR model, a functional response $y_i(t)$ is regressed upon the p scalar factors x_j, $j = 1,\ldots,p$ to produce functional effect estimates $\hat{\beta}_j(t)$ or $\hat{\beta}_{j,k}(t)$. That is,

$$y_{i,r}(t) = \beta_0(t) + \sum_{j=1}^{p} \beta_j(t)x_{i,j} + \sum_{j=1}^{p-1}\sum_{k=j+1}^{p} \beta_{j,k}(t)x_{i,j}x_{i,k} + \varepsilon_{i,r}(t), \quad i = 1,\ldots,n \tag{3}$$

for all time points $t \in \mathscr{T}$. In (3), $\varepsilon_{i,r}(t)$ is assumed to be drawn from a stochastic process with mean zero and covariance function $\Sigma(s,t)$, $s,t \in \mathscr{T}$ (Gertheiss et al. 2024).

Following Kokoszka and Reimherr (2017), the least squares estimator of $\beta(t)$ can be found by minimizing

$$\int_{\mathscr{T}} \sum_{i=1}^{n}\sum_{r=1}^{R} \left(y_{i,r}(t) - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j} - \sum_{j=1}^{p-1}\sum_{k=j+1}^{p} \beta_{j,k}x_{i,j}x_{i,k} \right)^2 dt \tag{4}$$

and thus for each t

$$\hat{\boldsymbol{\beta}}(t) = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\bar{\mathbf{y}}(t) \quad \text{where} \quad \bar{\mathbf{y}}(t) = (\bar{y_1}(t), \bar{y_2}(t), \ldots, \bar{y_n}(t))^{\mathrm{T}}. \tag{5}$$

The response in (3) is assumed to be a continuous function that can be evaluated at any time t. In practice however, $y(t)$ is typically available at a series of discrete observations which may be noisy, irregularly sampled or sparsely distributed (Gertheiss et al. 2024). In such cases, a common first step in building a functional meta-model is to smooth or interpolate the response data, as appropriate (Ramsay and Silverman 2005). However Kokoszka and Reimherr (2017) note that if the response function is defined over an equal grid with no missing observations, the raw data observations may be used directly.

There are limitations with the point-wise, or epoch-level, estimation of the coefficient functions $\boldsymbol{\beta}(t)$ in (5). For example, confidence intervals (CIs) calculated at each epoch do not necessarily translate to global confidence bands with nominal coverage probabilities for the entire function (Ramsay and Silverman 2005) and the smoothness of the resulting functional curve estimates for $\boldsymbol{\beta}(t)$ is also sensitive to the chosen number of epochs (Crainiceanu et al. 2024). Moreover, epoch-level estimation uses the raw observed data values and thus fails to capture the residual correlation expected in most function-on-scalar modeling applications (Crainiceanu et al. 2024).

To address these limitations, Crainiceanu et al. (2024) describe different modeling approaches in support of valid statistical inference with nominal coverage probabilities. In particular, Crainiceanu et al. (2024) propose a fast, scalable method based on bootstrapping smoothed estimates of $\boldsymbol{\beta}(t)$ from the epoch-level OLS regression. In this approach, bootstrap sampling with replacement is performed on the full functional response data, with epoch-level regression models derived for each individual bootstrap sample following (3) – (5). The resulting estimates of $\hat{\boldsymbol{\beta}}(t)$ are then smoothed across epochs to form the estimated coefficient functions for each bootstrap sample, with confidence intervals computed using standard error estimates from the bootstrapped results. Crainiceanu et al. (2024) demonstrate this approach provides plausible results while preserving the within-subject correlation structure of the data. Additionally the smoothing component may reduce sensitivity of results to the number of chosen epochs, while residual correlation is inherently accounted for as regression coefficients across epochs are related through residual covariance (Crainiceanu et al. 2024). Moreover, this scalable approach to achieving nominal coverage probabilities remains statistically efficient as the dimensions in the data grow, where other methods may become computationally infeasible.

3 COMBAT SCENARIO AND DATA

The combat engagement simulated for the purposes of this paper focused on a contested assault of a defensive position between two opposing forces (Red and Blue). The study objective was to determine the potential impacts on survivability and lethality of adding two prototype classes of remote autonomous system (RAS) capabilities to the assaulting (Blue) force structure. The potential confounding effects of four other combat-specific factors were also controlled for, namely *Force Structure*, *Tactics*, *Conditions* and *Environment*. Thus a total of six factors were included in the study. A 32 run, two-level, resolution VI fractional factorial design was used to model the main effects and all possible two-way interaction effects, as the quarter-fraction is only resolution *IV* and the two-way interaction effects would therefore be confounded with each other (Montgomery 2012).

The combat engagement was simulated using a closed-loop stochastic simulation model, with each design point replicated 151 times to capture the variability in the simulation results. Combat power over time formed the simulation response of interest, representing the percentage of initial forces remaining at a given point in time. In this way, combat power provides a proxy measure for the operational effectiveness of the Blue and Red forces.

4 META-MODELING OF COMBAT POWER AT END-OF-RUN

Simulation meta-models are derived separately in this section for the combat power of Blue and Red remaining at the end of the simulated engagement. These were derived using the *lm()* function in R, with variable selection performed using step-wise forwards and backwards elimination based on the values of the Akaike Information Criterion (AIC), a summary statistic for comparative model evaluation, with further backwards elimination applied based on the p-values (Weisberg 2005). Model validation is performed using model summary statistics combined with a visual comparison of the predicted versus observed data.

4.1 End-of-Run Meta-Model Results

The end-of-run meta-modeling results for the combat power of Blue and Red are shown in Table 1 and Table 2, respectively. In the meta-model for Blue, all six factors had a statistically significant effect (of different magnitudes) and all were involved in at least one significant two-factor interaction. Specifically for RAS capability, Table 1 shows the addition of either class significantly improves Blue survivability, however their combined effectiveness is diminished when operating together as evidenced by the negative interaction term in the meta-model. Thus, it is important to convey to military decision-makers that capabilities which interact, such as RAS here, should be viewed more as a portfolio instead of separately. *RAS Class I* did not have any other statistically significant interactions, however *RAS Class II* had a large amplifying interaction effect with *Force Structure*, and a smaller diminishing interaction with *Tactics*. The former observation is a classic example of a combat-multiplier, where the use of both *RAS Class II* and *Force Structure* is more effective than the sum of their individuals contributions, while the latter means that the effectiveness of *RAS Class II* will ultimately depend on the (unknown) *Tactics* chosen by Red (whereas *RAS Class I* does not suffer from this uncertainty). Again, both of these types of insights are particularly useful to military decision-makers. Finally, while *Force Structure* has a combat-multiplier effect with *RAS Class II*, it interacts negatively with the two non-Blue, non-Red factors (and more so with *Conditions* than *Environment*) and thus the overall effectiveness of *Force Structure* will be tempered by the specific situational elements of the scenario. Armed with these insights from the sensitivity analysis of the end-of-run meta-models, more informed decisions regarding the capabilities of Blue and their overall effectiveness can be made.

Table 1: Regression modeling results for combat power of Blue at the end of the simulated engagement. The estimated meta-model term effects are shown alongside the corresponding 95% CIs and p-values.

Meta-model term	$\hat{\boldsymbol{\beta}}$	95% CI	p-value
(Intercept)	81.09	80.30, 81.90	<0.001
RAS Class I	3.79	3.15, 4.43	<0.001
RAS Class II	4.85	3.95, 5.76	<0.001
Force Structure	8.02	7.11, 8.92	<0.001
Tactics	-1.01	-1.80,-0.23	0.014
Conditions	5.96	5.32, 6.60	<0.001
Environment	1.27	0.48, 2.05	0.003
RAS Class I * RAS Class II	-2.56	-3.47, -1.66	<0.001
RAS Class II * Force Structure	1.38	0.48, 2.29	0.005
RAS Class II * Tactics	-1.00	-1.90, -0.09	0.033
Force Structure * Conditions	-5.15	-6.05, -4.24	<0.001
Force Structure * Environment	-2.50	-3.41, -1.60	<0.001
Tactics * Environment	1.05	0.14, 1.95	0.026

R^2 = 0.987; Adjusted R^2 = 0.979; $\hat{\sigma}$ = 0.613; p-value <0.001; df = 12; Log-likelihood = -21.4; AIC = 70.8; BIC = 91.3; Deviance = 7.13; Residual df = 19; No. Obs. = 32.

Turning now to Table 2, in addition to improving Blue survivability, it can be seen that the addition of either RAS class also improves Blue lethality (as indicated by the negative main effect estimates), although in contrast to the meta-model for Blue, to a lesser extent (comparing the estimate magnitudes between tables) and without interaction. However, both RAS classes now significantly interact with *Force Structure*, but with a diminishing (not multiplying) effect. Neither *RAS Class I* or *RAS Class II* interact with the *Tactics* employed by Red (indeed *Tactics* had no effect at all on Red combat power), but *RAS Class II* joins *Force Structure* in interacting with the non-Blue, non-Red factor *Conditions*. Once again, these particular, and differing, insights can prove useful in military decision-making when balancing between measures of offensive and defensive effectiveness.

4.2 End-of-Run Meta-Model Validation

The simulation meta-models for Red and Blue end-of-run combat power are each statistically significant (p-value < 0.001) with the R^2 and adjusted-R^2 close to 1 for each model, indicating the models provide excellent fits to the data. This is supported by Figure 1 which shows strong agreement between the predictions from the end-of-run meta-models against the observed combat power values, particularly for Blue. There appears to be one potential outlier for the meta-model of Red combat power with Figure 1 showing the model appears to under-predict combat power for this design point.

5 FUNCTIONAL META-MODELING OF COMBAT POWER OVER TIME

This section extends the simulation meta-modeling results above to consider the evolution of factor effects on combat power over the course of the combat engagement. First, the distributions of average combat power over time for each design point in the study are explored for Blue and Red, averaged across all simulation replications. The scalable approach to function-on-scalar linear modeling described in Section 2.2 is then applied to model the effects of each factor on combat power over time. To preserve the desired structure in the data resulting from the experimental design, replications of each design point were resampled with replacement in the bootstrap procedure, in accordance with Kleijnen (2015). For each bootstrap resample, epoch-level OLS regression models are derived following (3) – (5) at intervals of 0.1 minutes. Note that

Table 2: Regression modeling results for combat power of Red at the end of the simulated engagement. The estimated meta-model term effects are shown alongside the corresponding 95% CIs and p-values.

Meta-model term	$\hat{\beta}$	95% CI	p-value
(Intercept)	21.60	20.10, 21.10	<0.001
RAS Class I	-0.76	-1.18, -0.33	0.001
RAS Class II	-1.38	-1.90, -0.86	<0.001
Force Structure	-2.87	-3.47, -2.27	<0.001
Conditions	-3.14	-3.74, -2.54	<0.001
Environment	-0.75	-1.17, -0.32	0.001
RAS Class I * Force Structure	0.73	0.13, 1.33	0.020
RAS Class II * Force Structure	0.92	0.32, 1.52	0.005
RAS Class II * Conditions	1.04	0.44, 1.64	0.002
Force Structure * Conditions	1.71	1.11, 2.31	<0.001
Conditions * Environment	1.02	0.42, 1.62	0.002

R^2 = 0.919; Adjusted R^2 = 0.881; $\hat{\sigma}$ = 0.408; p-value <0.001; df = 10; Log-likelihood = -10.0; AIC = 44.0; BIC = 61.6; Deviance = 3.50; Residual df = 21; No. Obs. = 32.

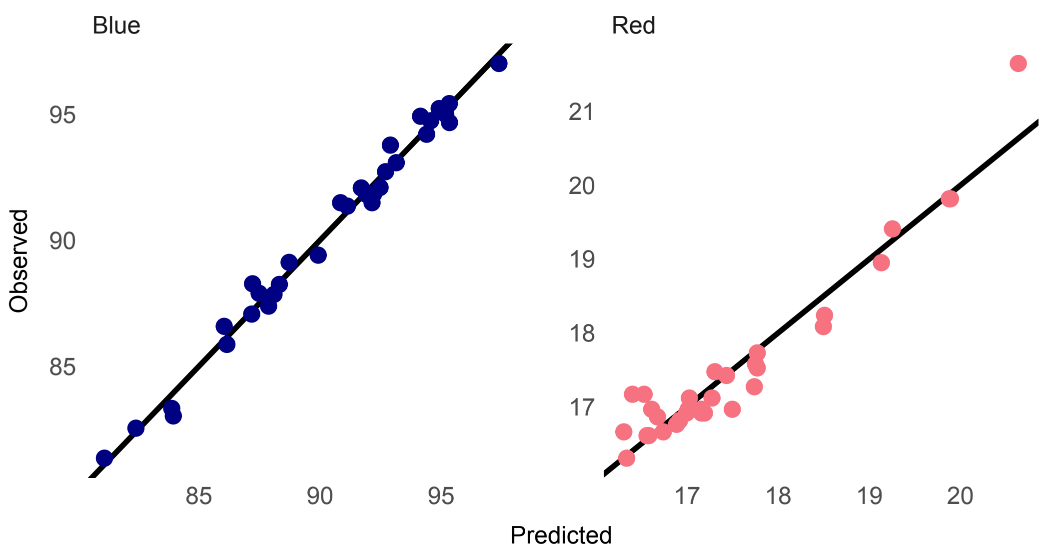

Figure 1: Predicted versus actual plots for the meta-models of end-of-run combat power for Blue and Red. The diagonal black line represents the expected values if the model perfectly predicted the observed data.

as the combat power data were available at an equal grid of points with no missing values, the raw data observations are used directly in the functional meta-modeling without prior smoothing.

Effect estimates are then smoothed across epochs using LOWESS smoothing via the *tf_smooth()* function in the *tf* library in R (Cleveland 1979). Smoothing parameter values of 0.05 for Blue and 0.02 for Red were found to satisfactorily capture the trend in the effect estimates without over-fitting the data. These are small parameter values, however larger smoothing windows failed to capture the variation in peaks and troughs that may be expected with typical combat simulation where events can occur in quick succession leading to rapid changes in factor effects. Thus larger values of the smoothing parameter resulted in non-homogeneous, non-symmetric patterns in the residual plots over time. Conversely, smaller values of the smoothing parameter tended to over-fit the data, providing an unsatisfactorily noisy representation of the trend in effect estimates. The smoothing parameters of 0.02 and 0.05, corresponding to respective smoothing windows of approximately 48 seconds and 2 minutes around each data point, were used across all effect curves at each bootstrap resample, however it should be noted that results may be improved by optimizing the parameter selection for each effect curve and bootstrap resample individually however this was beyond the scope of this research.

The resulting functional meta-models are then validated by examining $R^2(t)$, a function analogous to the R^2 statistic from multivariate analysis but with an added dependence on time (Ramsay and Silverman 2005). This is supplemented with an analysis of residual distributions over time by design point.

5.1 Functional Meta-Modeling Results

Figure 2 shows the average combat power over time for Blue and Red, with each line corresponding to a distinct design point representing a unique combination of the six study factors. From Figure 2, Blue can be seen to typically dominate the engagement, retaining a high percentage of combat power relative to Red across all considered time points. There is some variability in combat power over time between design points however, most notably for Blue, providing some evidence of non-constant factor effects over time. Combat power also remained relatively constant during the earliest stages of the engagement ($t < 8$ minutes for Blue and $t < 6$ minutes for Red) at which points the two forces were advancing to the main engagement. Similarly, combat power was also constant after approximately $t = 30$ minutes, following the conclusion of the main engagement.

The functional meta-modeling results for Blue and Red combat power over time are each shown in Figure 3 and provide strong evidence of heterogeneous factor effects over the course of the simulated engagement. Specifically, Figure 3 shows the main effects of *RAS Class I* and *RAS Class II* on Blue combat power are not constant, with the effect curves not statistically significant until approximately $t = 25$ minutes, at which point there was a sharp increase in the effect of each RAS class. The effect curves then remained relatively constant for the remainder of the engagement. Similarly, the interaction between the two RAS classes also emerges as statistically significant at approximately this time point. A review of the events occurring within the simulation at $t = 25$ minutes indicated that these effect changes were in response to a significant acute event in the simulation, namely an ambush of Blue by Red. This suggests that the effectiveness of the added RAS capability was tied to this specific event. In contrast, the effects of *Force Structure* and *Conditions* each increased steadily over the majority of the engagement, indicating

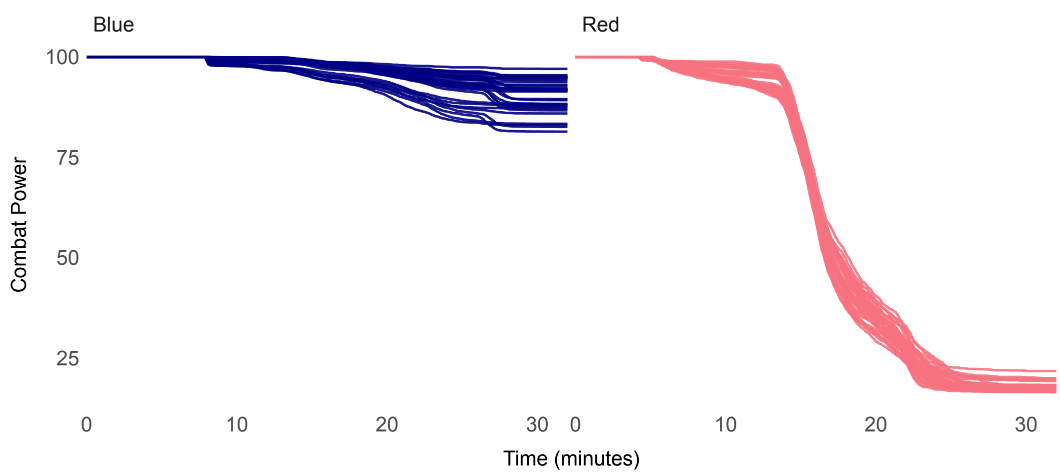

Figure 2: Average combat power over time response curves by design point for Blue and Red. Each line in the figure corresponds to a unique combination of the six study factors.

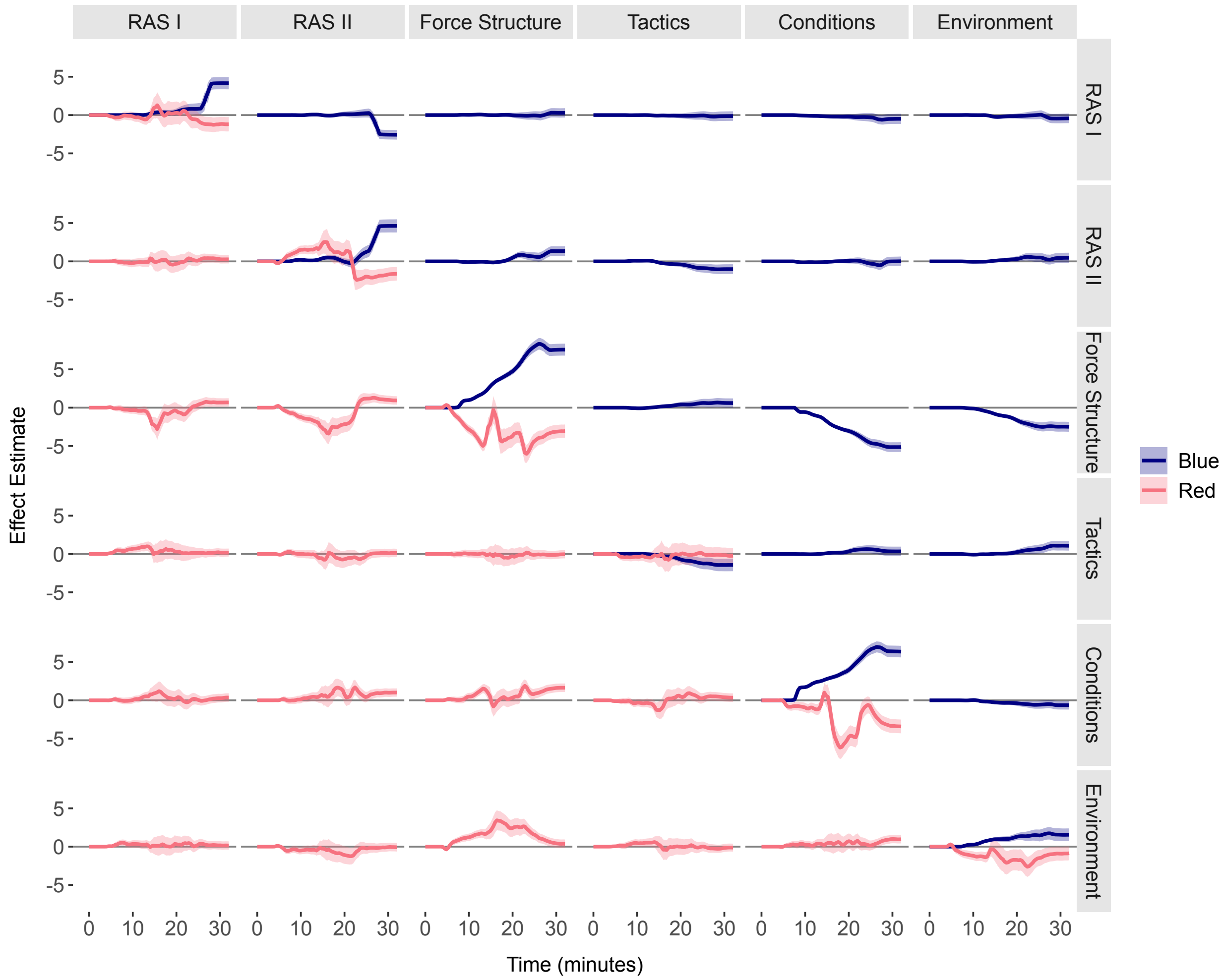

Figure 3: Functional meta-model results for Blue and Red combat power over time showing the estimated factor effects (solid line) and corresponding 95% CIs (shaded ribbon). The main effect of each factor is shown along the diagonal with the off-diagonal elements presenting the interaction effects between factors in the corresponding row and column. Here, color is used to distinguish between the functional meta-modeling results for Blue and Red.

these factors had beneficial effects on Blue combat power throughout the scenario and were not restricted to a specific event. Thus the functional meta-models enable an enriched characterization of the factor effects over time, providing analysts with insight into how and where effects are occurring during the combat engagement. This level of insight is not possible from the end-of-run meta-models alone.

Similar observations can be made for the combat power of Red (Figure 3) however with critical additional insight into the changing nature of effects over the combat engagement. Specifically the effect of *RAS Class II* changes sign from positive to negative over the course of the combat engagement. Thus, relative to not including the capability, the addition of *RAS Class II* to the Blue side had a positive effect on Red combat power during the early stages of the engagement, progressing to a negative effect during the later stages. The direction of the interaction effect between *RAS Class II* and *Force Structure* also changes from a significant negative effect at approximately $t = 20$ minutes to a significant positive effect from approximately $t = 25$ minutes onward. This insight into the reversal of effects is not possible from the end-of-run simulation meta-modeling results (Table 2) which showed only that *RAS Class II* had a significantly negative effect on Red combat power while the *RAS Class II* had a significant positive interaction with *Force Structure* at the end of the simulation.

Moreover, Figure 3 shows a number of main and interaction effects which are either non-significant or near-zero at end-of-run, however have a larger significant effect during the simulation. One clear example of this is the interaction effect between *Force Structure* and *Environment* which has a significant positive effect on Red combat power across most time points, however decreases to zero at the end of the engagement. Interaction effects such as this may be expected in combat engagements however the end-of-run simulation meta-models only capture effects at the end of the engagement, this interaction effect is completely missed by these meta-models with the regression results in Table 2 indicating no significant effect between *Force Structure* and *Environment*. The additional temporal insight from the functional meta-models therefore enables an enriched understanding of the nature of factor effects during the combat engagement, which the end-of-run meta-models may fail to capture, in support of better informed decision-making.

Lastly, functional linear meta-modeling may also lend an additional degree of confidence to the end-of-run results for effects which remain constant over the simulated scenario. For example, the interaction effect between *Force Structure* and *Tactics* was not found to be statistically significant in the end-of-run meta-models for Blue or Red (Table 1 and Table 2, respectively), with the functional meta-modeling results (Figure 3) also showing these interaction effects remained insignificant across all time points. Thus the functional results provide an additional measure of trust in the end-of-run modeling results that there are no hidden effects which could otherwise be missed for these terms.

5.2 Functional Meta-Model Validation

Figure 4 shows $R^2(t)$ for the functional meta-models of combat power over time for Blue and Red, providing a measure of explanatory power. The results indicate that the meta-model for Blue provides an excellent fit to the data across all time points and may potentially be over-fitting the data. This may be due in part to the inclusion of all main effects and all possible two-way interactions in the model however, due to the sheer number of regression models fit over all time points, variable selection quickly becomes complex in a functional meta-modeling context. In contrast, Figure 4 shows some slight variation in model fit over time for Red, with $R^2(t)$ ranging from nearly 1 at most time points and dropping to 0.822 at $t = 15.7$ minutes. The timing of this drop coincides with the greatest decrease in combat power for Red (Figure 2), with further analysis of the data indicating this time also had the greatest variability in combat power across all design points for Red. Despite this decrease, the R^2 value was still close to 1 indicating the fitted functional meta-model provided a strong fit to the data.

Figure 5 shows the residuals from the functional meta-models of combat power over time for Blue and Red. It can be seen that variance in the residuals is not constant over time, reflecting the variability in combat power at different points in time that was observed in Figure 2. The residuals appear to be zero

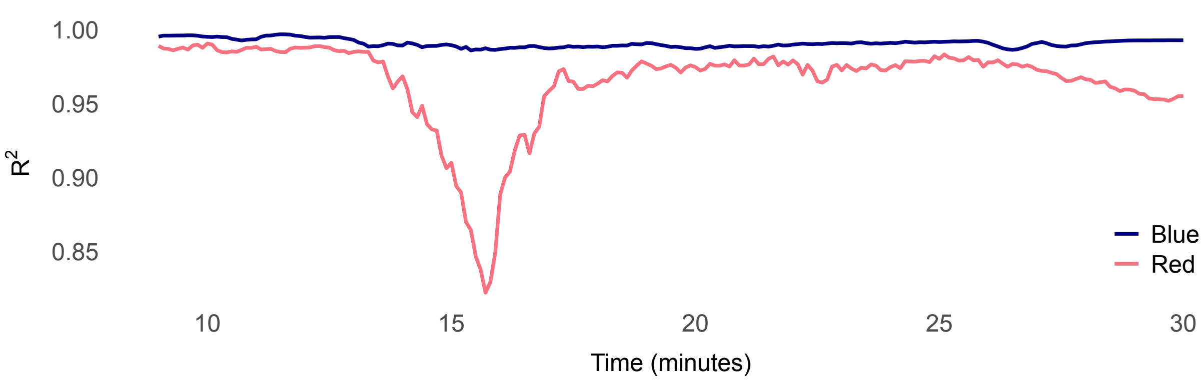

Figure 4: $R^2(t)$ for the functional meta-models of combat power over time for Blue and Red.

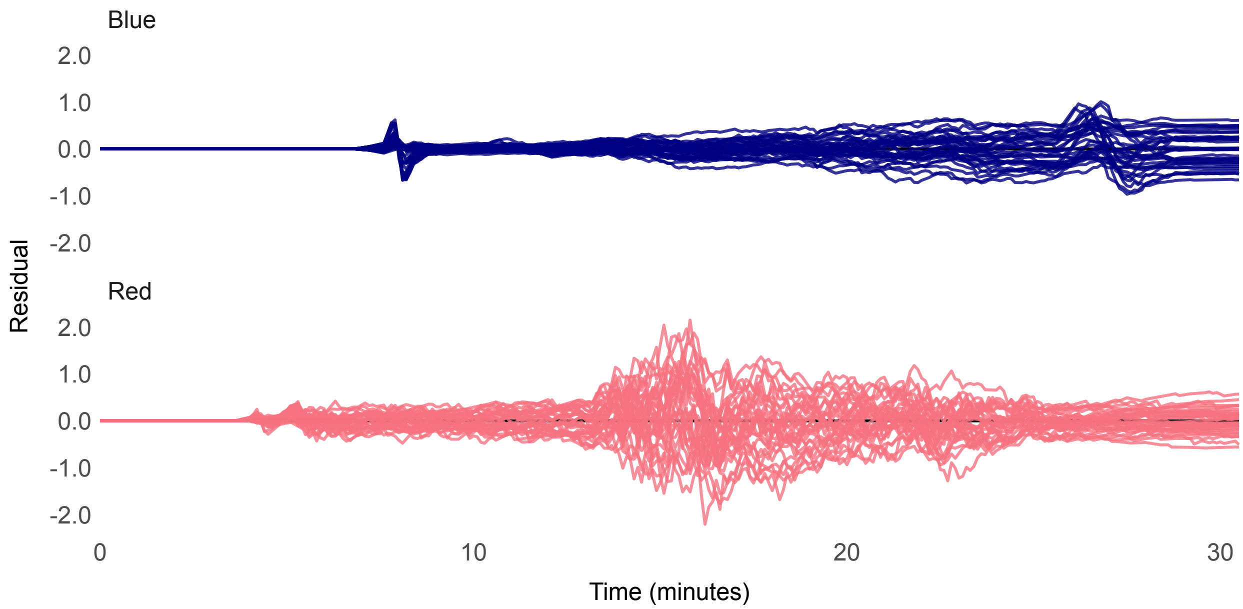

Figure 5: Residuals over time from the functional meta-models of combat power over time for Blue and Red, respectively. Each line in the figure corresponds to a different design point in the study.

until approximately $t = 8$ minutes for Blue and $t = 6$ minutes for Red, in direct response to the constant combat power preceding these time points (Figure 2). The residuals also exhibit an unusual 'saw-tooth' pattern at approximately $t = 8$ and and $t = 27$ minutes for Blue, with a similar, albeit less pronounced, pattern at approximately $t = 5$ minutes for Red. This is likely attributable to the smoothing step in deriving the functional meta-models, with data points in a neighborhood around each observation having a direct influence on the shape of the smoothed curve at that observation. The constant combat power during the early and late stages of the simulation may have therefore compromised the quality of the smoothed results at adjacent time points.

The magnitude of the residuals in Figure 5 are no greater than 1 and 2 for the functional meta-models of Blue and Red, respectively, with the residuals appearing reasonably symmetric around zero indicating a lack of bias. Thus despite the drop in model quality at $t = 15.7$ minutes and the potentially reduced performance of the smoothing estimates in the tails of the distributions, the functional meta-models appear to provide a strong fit to the combat power over time data for both Blue and Red.

6 SUMMARY AND FUTURE RESEARCH

Traditional sensitivity analyses employ meta-models fit to end-of-run simulation responses, yet simulations can be asked to output a range of intermediate variable states during its course of execution. This paper sought to examine if meta-models fit to this functional response of a combat simulation could enable an enhanced sensitivity analysis and provide a richer characterization of the effects of military capabilities on mission effectiveness, thereby improving analytical support to military decision-making. A traditional two-level fractional factorial experimental design, with replication, was used to fit linear response surface meta-models relating six factors of interest to the percentage of forces remaining at different time points during a combat engagement between Blue and Red. This functional meta-modeling of combat power over time was then contrasted with the typical end-of-run results to discover what, if any, additional insight might emerge. Of particular interest from the case study examined was the identification of acute effects that occurred only at specific time points, the existence of an intermediate sign-change of effect (the main effect of one factor and a two-way interaction affecting Red's combat power over time) as well as a factor having a statistically significant effect during the combat engagement, but not at it's conclusion. These insights were not possible from the end-of-run sensitivity analysis alone, with the functional meta-models therefore providing an additional depth of insight into the nature of factor effects over the course of the combat engagement.

It should be noted that the functional linear modeling method used here is just one of several implementations and was selected because it is both fast, scalable and explicitly accounts for residual correlation across time. Future research will explore other implementations to determine the modeling approach best suited to functional data from combat simulation as well as sensitivity of the functional modeling results to choice of smoothing parameters. Furthermore, to better reflect the true nature of typical combat measures (e.g., count or binary responses), extending this linear functional meta-modeling to generalized linear modeling, similar to James (2002) but with a functional response and scalar predictors, will also be undertaken. Variable selection procedures for functional modeling applications will also be investigated to mitigate the potential for over-fitting. Lastly, future research will also explore the extent to which anomaly detection procedures can be applied to functional response data from combat simulations, particularly in the context of detecting anomalous simulation replications.

REFERENCES

Balogh, I., and G. Harless. 2003. "An Overview of the COMBATXXI Simulation Model: A Model for the Analysis of Land and Amphibious Warfare". In *Proceedings of 71st Military Operations Research Society Symposium*. June, Virginia, USA.

Cleveland, W. S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots". *Journal of the American Statistical Association* 74(368):829–836.

Crainiceanu, C., J. Goldsmith, A. Leroux, and E. Cui. 2024. *Functional Data Analysis With R*. Florida, USA: CRC Press.

Dunn, P., and G. Smyth. 2018. *Generalized Linear Models with Examples in R*. New York, USA: Springer New York.

Fedorov, V. V. 1972. *Theory of Optimal Experiments*. New York, USA: Academic Press.

Gertheiss, J., D. Rügamer, B. X. Liew, and S. Greven. 2024. "Functional Data Analysis: An Introduction and Recent Developments". *Biometrical Journal* 66(7):e202300363.

Gill, A. 2021. "Heteroscedasticity and Correlation in Linear Regression". In *Proceedings of MODSIM2021, 24th International Conference on Modelling and Simulation*. December 5th-10th, Sydney, Australia.

Gill, A., D. Grieger, M. Wong, and W. Chau. 2018. "Combat Simulation Analytics: Regression Analysis, Multiple Comparisons and Ranking Sensitivity". In *2018 Winter Simulation Conference (WSC)*, 3789–3800 https://doi.org/10.1109/WSC.2018.8632245.

James, G. M. 2002. "Generalized Linear Models with Functional Predictors". *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 64(3):411–432.

Kleijnen, J. 2015. *Design and Analysis of Simulation Experiments*. 2nd ed. Cham, Switzerland: Springer Cham.

Kokoszka, P., and M. Reimherr. 2017. *Introduction to Functional Data Analysis*. New York, USA: Chapman and Hall/CRC.

Lauren, M., and R. Stephen. 2002. "Map-Aware Non-uniform Automata (MANA) – A New Zealand Approach to Scenario Modelling". *Journal of Battlefield Technology* 5:27–31.

Montgomery, D. 2012. *Design and Analysis of Experiments*. 8th ed. New Jersey, USA: John Wiley & Sons, Inc.

Ong, J., and M. F. Ling. 2002. "Using the Joint Integrated Contingency Model for Campaign Analysis". Technical Report DSTO-TR-1307, DSTO Electronics and Surveillance Research Laboratory.

Ramsay, J., and B. Silverman. 2005. *Functional Data Analysis*. 2nd ed. New York, USA: Springer New York.

Sanchez, S. M., T. W. Lucas, P. J. Sanchez, C. J. Nannini, and H. Wan. 2012. "Designs for Large-Scale Simulation Experiments, with Applications to Defense and Homeland Security". In *Design and Analysis of Experiments: Special Designs and Applications*, edited by K. Hinkelmann, 413–441. New Jersey, USA: John Wiley & Sons, Ltd.

Santner, T. J., B. J. Williams, and W. I. Notz. 2003. *The Design and Analysis of Computer Experiments*. New York, USA: Springer New York.

Scherer, M. E. 2021. *Analysis of Functional Responses in Experimental Design*. Master's thesis, Air Force Institute of Technology, Ohio, USA. https://scholar.afit.edu/etd/4933, accessed 20th June 2024.

Seymour, C. N. 2014. *Capturing the Full Potential of the Synthetic Theater Operations Research Model (STORM)*. Ph. D. thesis, Monterey, California: Naval Postgraduate School.

Weisberg, S. 2005. *Applied Linear Regression*. New Jersey, USA: John Wiley & Sons, Inc.

Zeh, J., B. Birkmire, P. Clive, J. Johnson, A. Krisby, J. Marjamaa, *et al*. 2014. "Advanced Framework for Simulation, Integration, and Modeling (AFSIM) Version 1.8 Overview". Technical Report 88ABW-2014-5041, Air Force Research Laboratory, Aerospace Systems.

AUTHOR BIOGRAPHIES

LISA BLUMSON is a senior researcher and data analytics lead for combat simulation studies at the Defence Science and Technology Group in Australia. She holds a PhD in Mathematics from the University of South Australia and has a background in applying statistical modeling, data analysis and optimization techniques across a diverse range of research applications. Her email address is lisa.blumson1@defence.gov.au.

CHARLIE PETER is a data analyst working in combat simulation at the Defence Science and Technology Group in Australia. He holds a BSc (Hons) in Applied Mathematics from the University of Technology in Sydney, with his work focusing on operations research. His email address is charlie.peter@defence.gov.au.

ANDREW GILL is a senior operations analyst at the Defence Science and Technology Group in Australia. He holds a PhD in Applied Mathematics from James Cook University and a MSc in Operations Research and Statistics from the University of New South Wales. His research interest is on the design and analysis of simulation experiments. His e-mail address is andrew.gill@defence.gov.au.