SCOPE: An Integrated Early Warning System for Combat Casualty Prediction

J.K.R. Tiong [a], **B.A. Chiera** [a]

[a] *UniSA STEM, University of South Australia, Adelaide, South Australia, 5000*
Email: junia.tiong@unisa.edu.au

Abstract: Combat simulations generate large volumes of time-stamped event series that record tactical activity and mission outcomes. Historically, analysis has relied on retrospective evaluation of end-of-run metrics, which limits operational foresight. The ability to identify signals of increasing vulnerability to adverse outcomes during the course of engagement offers a significant advantage by supporting earlier intervention and more informed decision-making. This study investigates whether segmenting combat event series into critical windows of activity can provide early indications of mission risk before terminal outcomes are realised.

To achieve this objective, motif bigrams are introduced as compact and interpretable features that represent short-range tactical transitions within event series. These features underpin the Score-based Combat Operational Predictive Event (SCOPE) system, an integrated early warning framework that predicts mission outcomes by continuously monitoring combat dynamics. SCOPE adopts a layered modelling strategy that combines machine learning and regression-based methods to balance predictive accuracy, interpretability, and robustness across scenarios. The framework further incorporates an audit layer to validate predictor influence across replications, ensuring transparency and accountability in the identification of key tactical drivers of mission risk.

A high-fidelity combat simulator was used to generate detailed operational data across three scenarios, each with 201 replications. Each replication produced a time-ordered event series composed of Movement, Detection, Shot, and Kill events, recorded at irregular time intervals, between opposing force teams *Blue* and *Red*. These series were represented using 16 motif bigram types preserving local structure within fixed-length time windows. To provide a continuous and sensitive measure of Blue force attrition, a *Casualty Score* was developed by weighting losses across key entity types, allowing more nuanced outcome assessment than categorical mission end states.

Predictive modelling was undertaken to estimate the Casualty Score using motif bigram counts. Random Forest, Generalised Additive Models, and Multivariate Adaptive Regression Splines were evaluated, each contributing distinct strengths in variable selection, interpretability, and non-linear modelling. A fixed feature set of eight bigrams, selected via Random Forest importance rankings, was used to enhance stability and comparability across scenarios. While no single method fully satisfied the requirements for accuracy, clarity, and adaptability, their complementary strengths were integrated into a layered modelling framework that underpins SCOPE.

Results demonstrate that motif-based features linked to the Casualty Score enable accurate and interpretable forecasts of combat outcomes across scenarios. The layered design of SCOPE enables both robust prediction and operational insight, with an interactive interface that provides predicted risk scores, interpretive plots, and real-time comparisons against observed losses. By integrating predictive modelling with structured interpretability, SCOPE supports adaptive monitoring of combat progression and informed tactical decision-making. The approach presented here offers a practical foundation for developing early warning systems in combat simulations, with potential to extend to real-world operational contexts where timely recognition of emerging risk is critical.

Keywords: *Combat simulation, early warning, motif bigrams, casualty prediction, event sequence analysis*

1. INTRODUCTION

Combat simulations generate extensive time-stamped event data that capture tactical activities and mission outcomes. While such data have traditionally supported retrospective analyses of performance and system behaviour (Schubert et al., 2015), there is a growing interest in identifying early indicators of mission risk during the course of an engagement. Early Warning Systems (EWS) are designed to detect and communicate impending threats, enabling timely interventions to mitigate adverse outcomes. These systems have been successfully applied across multiple domains, including ecological and environmental contexts (Scheffer et al., 2009), agricultural settings (Brown et al., 2022), and military operations (Mathais et al., 2025).

In the military context, the ability to detect signs of growing vulnerability during key stages of combat offers a notable advantage, enabling early intervention and better-informed decision making (Chmielewski et al., 2020; Mathais et al., 2025). Traditional combat modelling generally focuses on summarising outcomes such as kill ratios, resource use, or mission success at the end of a simulation run and pays little attention to how combat interactions unfold over time (Masterson, 2004). Early warning systems have been demonstrated in industrial settings, for example in the automotive sector where simulation-based approaches are used to detect risks during operations (Hotz et al., 2006), but their application to combat behaviour modelling remains largely unexplored. This study adopts an alternative approach by segmenting time-ordered event series data to gain early insight into adverse mission outcomes. By identifying meaningful shifts in combat behaviour as it unfolds, this method supports timely interpretation of mission progression for improved decision making.

Building upon these principles, this study introduces the Score-based Combat Operational Predictive Event (SCOPE) system, an integrated early warning framework that predicts mission outcomes by continuously monitoring combat dynamics. SCOPE employs motif bigrams as compact, interpretable features representing short-term tactical transitions within event series. These features provide a stable foundation for prediction, where traditional time series and regression methods often fail (Lin et al., 2002). SCOPE adopts a layered modelling strategy combining machine learning and regression methods to balance predictive accuracy, interpretability, and robustness across scenarios. Additionally, SCOPE incorporates an audit layer to validate predictor influence across replications, ensuring transparency and accountability in the identification of key tactical drivers of mission risk.

This paper is structured as follows. Section 2 presents the combat simulator and event data series and a proposed metric for combat attrition. Section 3 details the development and evaluation of candidate models for core prediction, including interpretation and audit methods. Section 4 introduces the integrated early warning system, highlighting modelling layers and user interface. Section 5 discusses broader implications and future directions.

2. COMBAT SIMULATOR

This study relies on a high-fidelity, closed-loop combat simulator that generates detailed operational data through realistic and repeatable mission simulations, recording events in discrete time. The simulator is designed to support decision-making across a range of operational conditions (Chau et al., 2017). A scenario is defined by a specific set of input parameters that configure the combat environment. These parameters represent competing options relevant to tactical planning or what-if analysis, allowing different aspects of mission design to be explored systematically. The simulator models engagements between two opposing forces, focusing here on the assaulting force called the Blue team. The simulator records outcomes through categorical end-of-run metrics such as Mission Success or Mission Failure, alongside other discrete, binary, and continuous measures. High-level metrics summarise engagements while providing limited insight into attrition dynamics.

2.1. The Event Data Series and Casualty Score Metric

The combat data under consideration has three scenarios, A, B and C, used to represent distinct combinations of simulator input options. Each scenario was replicated 201 times using a fixed sequence of non-repeating pseudorandom simulation seeds. This replication process is designed to reflect the stochastic nature of combat and preserves within-scenario variability while enabling consistent comparison across different scenarios. Each replication yields a chronological sequence of time-stamped combat events with four event types: Movement (M), Detection (D), Shot (S), and Kill (K). Events are recorded as they occur, resulting in non-homogeneous temporal spacing leading to temporal sparsity, challenging standard time series methods.

To describe combat event series, this current analysis uses motif bigrams – pairs of consecutive events drawn from the set $\{M, D, S, K\}$ that capture short-range dependencies in tactical behaviour. For example, $\{DS\}$ indicates a detection was immediately followed by a shot, while $\{MK\}$ reflects a kill that occurred after movement. The length-2 structure was selected based on a Minimum Description Length (MDL) test (Grünwald, 2007), which identified bigrams as providing the best balance between model complexity and informational value. Prior work also supports this approach, which employed a first-order Markov model to detect anomalies (Tiong et al., 2023). Bigrams are advantageous as they capture short-range dependencies, to enable early detection of shifts in engagement patterns, coordination, or vulnerability within evolving combat sequences. All 16 possible bigram combinations are included as predictor variables for SCOPE:

$$\{MM, MD, MS, MK, DM, DD, DS, DK, SM, SD, SS, SK, KM, KD, KS, KK\}. \tag{1}$$

The bigrams are counted within fixed-length time windows over the event series to preserve local temporal structure, captures short-term behavioural patterns. These counts serve as inputs into SCOPE to predict mission outcomes based on recent patterns in the simulated event series.

This study introduces a new output metric, the continuous-valued Casualty Score, to provide a more detailed assessment of mission outcomes. While the combat simulator produces categorical end-of-run metrics, these summarise outcomes coarsely and lack granularity. The Casualty Score addresses this limitation by integrating losses across tanks, infantry fighting vehicles (IFVs), and infantry through a weighted formulation that reflects the simulator's terminating conditions and the relative strategic value of each unit type. Specifically, termination is triggered once any one of the following thresholds is reached: the loss of all three tanks, one-third of the IFVs, or one-third of the infantry force. To capture this within a continuous measure, the Casualty Score is expressed as:

$$Casualty\ Score = \omega_T N_{Tank} + \omega_I N_{IFVs} + \omega_F N_{Infantry}, \qquad \omega_T > \omega_I > \omega_F \tag{2}$$

where N_{Tank} , N_{IFVs} , $N_{Infantry}$ denote the number of neutralised units, and the weights $\omega_T, \omega_I, \omega_F$, the exact values of which will be specific to scenarios, are ordered to reflect their relative importance, with weights bounded to ensure a continuous score that approximates termination logic. This continuous metric quantifies Blue team attrition and supports finer comparisons across simulation replications than categorical outcomes alone.

3. MODEL DEVELOPMENT AND EVALUATION

This section details the predictive models used to estimate the Casualty Score – using irregular sequences of combat events. To support the SCOPE framework, motif bigrams are extracted from the event series and used as predictors, capturing local combat dynamics while accommodating the sparse and non-uniform nature of the data. The modelling pipeline evaluates three approaches: (1) Random Forest (RF); (2) Generalised Additive Models (GAM); and (3) Multivariate Adaptive Regression Splines (MARS) with each approach offering complementary strengths in predictive accuracy, interpretability, and flexibility.

3.1. SCOPE Framework

The combat event series presents two main challenges: temporally non-homogenous events and short-range dependencies embedded within motif bigram frequencies. These characteristics violate the assumptions of traditional linear models, such as fixed time intervals and independent predictors. To overcome these limitations, flexible, non-parametric models are adopted to accommodate structural irregularity and complex local dynamics.

A Random Forest model was initially considered due to its robustness in high-dimensional settings and its ability to capture non-linear interactions without prior specification however its ensemble structure limits interpretability, making it difficult to isolate the contribution of individual variables. To address this limitation, Generalised Additive Models (GAMs) were explored. The GAM model smooths additive effects for each predictor and is well suited to visualising changes in the risk of casualties as motif frequency increases. However, the additive form of a GAM restricts the modelling of interactions or abrupt transitions unless explicitly incorporated. Multivariate Adaptive Regression Splines (MARS) offered a flexible alternative for the objective here by constructing piecewise linear functions between segments of the predictor space and automatically selecting relevant interactions. MARS was trialled in both scenario-specific and pooled configurations. In the pooled configuration, data from all 201 replications across Scenarios A, B, and C were combined into a single training set. This approach aimed to capture broader regularities in the relationship between motif bigrams and values taken by the *Casualty Score* metric, while remaining sensitive to contextual variation across different combat settings.

3.2. Variable Selection and Final Feature Set for SCOPE

Initial modelling began with the sixteen motif bigrams listed in (1). These were treated as predictor variables, collectively forming the feature set. A Random Forest (RF) model was fitted to the complete dataset of 201 replications for Scenarios A, B and C to predict the *Casualty Score* and quantify attrition of the defensive Blue team. Predictor importance was assessed using the percentage increase in mean squared error *(%IncMSE)* statistic, which measures the drop in model performance when each variable is permuted. Results showed that predictive power was concentrated in a core subset of eight bigrams. Variables beyond this subset contributed marginally to accuracy while increasing model complexity and computational cost. Accordingly, the feature set was reduced to the top eight bigrams based on their *%IncMSE* rankings, with consistent importance observed across scenarios.

Figure 1 presents the *%IncMSE* scores for all sixteen motif bigrams. The top eight, indicated in blue, were selected for all subsequent modelling based on their stability, relevance across scenarios, and predictive contribution. Of the top eight variables in Scenario A, six also appeared in the top eight for Scenario B, while all eight from Scenario C matched the top eight in Scenario A, albeit with different rankings. This fixed feature set was applied consistently across models to support comparability, enhance interpretability and reduce noise from scenario-specific artefacts such as overfitting to rare event patterns.

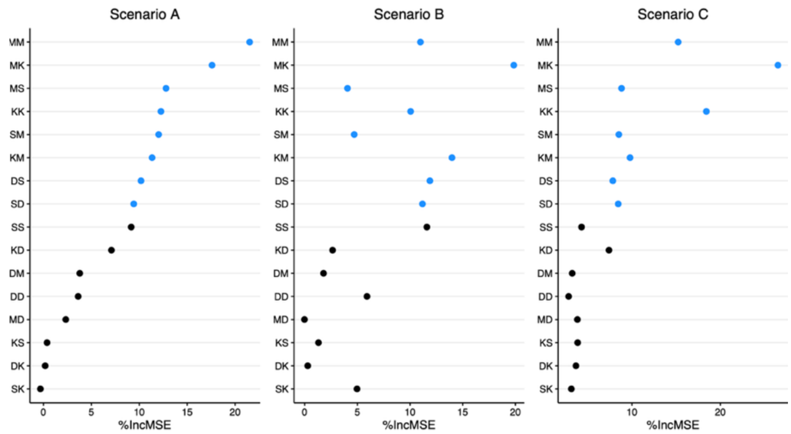

Figure 1: Random Forest variable importance *(%IncMSE)* for 16 bigram predictors across Scenarios A, B and C.

3.3. Comparative Model Performance

Following variable selection for SCOPE, the Random Forest model was re-run using the top eight predictors for direct comparison with GAM, MARS, and pooled MARS. All models in this subsection used the same reduced feature set. Evaluation was conducted across Scenarios A, B, and C using a 70:30 replication-level split to preserve scenario integrity and prevent overlap between training and test sets. This split was held constant across models to ensure comparability. To mitigate overfitting, the final models also restricted predictors to the reduced set, applied cross-validation during tuning, and incorporated bootstrapped confidence intervals. Performance was assessed using generalised R^2, RMSE, MAE (Table 1).

Although raw performance values were broadly similar across models, pooled MARS was selected for its stability and generalisability across scenarios while maintaining both accuracy and interpretability. Scenario-level Friedman tests demonstrated a significant difference between models in Scenario A ($p = 0.029$), but no significant differences in Scenarios B or C, highlighting the relative similarity of results. Pooled comparisons of replication-level R^2 across all scenarios using Wilcoxon signed-rank tests further demonstrated that pooled MARS consistently outperformed GAM, scenario-specific MARS, and both RF variants ($p < 0.05$). Thus, pooled MARS not only matched competing methods in raw accuracy but also offered a statistically defensible advantage when considered across scenarios. RF performed strongly, particularly in Scenarios B and C, but lacked interpretability. Scenario-specific MARS models adapted to local variation but produced greater variability. GAM generated interpretable smooth fits, showing how casualty risk shifted with motif frequency, but its additive form limited the capture of abrupt thresholds or interactions.

Scenarios / Models	A			B			C			Pooled		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
RF (16 predictors)	0.63	0.81	0.67	0.58	0.81	0.65	0.69	0.80	0.66			
RF (8 predictors)	0.60	0.83	0.68	0.59	0.80	0.64	0.68	0.79	0.66			
GAM	0.64	0.76	0.62	0.62	0.74	0.61	0.68	0.74	0.60			
MARS	0.60	0.80	0.64	0.55	0.81	0.66	0.67	0.76	0.61			
Pooled MARS										0.66	0.754	0.61

Table 1: Comparative model performance across Scenarios A, B, and C.

Motif bigrams offer a compact and interpretable representation of tactical sequences to support predictive modelling. Limitations include ignoring temporal spacing between events, thus identical bigrams may have different implications depending on their occurrence rate, uneven representation of bigrams that can bias models, and restriction to consecutive-event transitions, which overlooks broader sequence context that may influence casualty outcomes. Despite these constraints, pooled MARS effectively leverages the available information to provide stable and generalisable predictions. Each model offers distinct strengths and limitations. RF delivers strong predictive accuracy but limited transparency. GAM supports interpretable fits but struggles with sharp changes and interactions. Scenario-specific MARS adapts to local structure but varies in accuracy. By contrast, pooled MARS balances predictive accuracy, interpretability, and generalisability across scenarios, justifying its role as the core predictive engine for SCOPE. GAM overlays this foundation to visualise motif to casualty relationships, while RF supports variable selection via importance ranking. Together, the models operate sequentially, combining predictive strength, interpretability, and diagnostic insight, forming the basis of the integrated SCOPE framework introduced in Section 4.

4. SCOPE: SYSTEM ARCHITECTURE AND INTERFACE OVERVIEW

This section presents the Score-based Combat Operational Predictive Event (SCOPE) system. SCOPE functions as an integrated early warning capability for combat simulations, detecting signs of emerging mission risk by analysing short-range patterns in event sequences. These patterns are transformed into motif bigrams within sliding time windows and linked to the Blue team *Casualty Score* metric. The following subsections describe the structure of the modelling framework and its integration into an interactive interface for real-time risk assessment.

4.1. System Integration

At the core of SCOPE is a layered modelling structure designed to integrate prediction, interpretation, and diagnostic audit. Sequences of tactical events are transformed into motif bigrams within fixed-length sliding windows, capturing local behavioural dynamics that often precede mission degradation. These patterns are then linked to the continuous *Casualty Score* (2). SCOPE adopts a layered modelling structure that integrates prediction, interpretation, and diagnostic audit. At its centre is a pooled MARS model that estimates nonlinear relationships between motif frequencies and the *Casualty Score*. This predictive layer is supported by two interpretive components. A generalised additive model (GAM) is used to estimate smoothing effects for each motif, offering scenario-specific insight into how changes in local event series patterns contribute to predicted casualty levels. A Random Forest model is also included to perform a variable audit, quantifying the importance of each predictor variable across replications. This audit layer supports transparency by identifying motifs with consistent influence, as well as those whose contribution varies by operational context.

This combination of models enables SCOPE to provide accurate and interpretable predictions of casualty outcomes. Predictive signals embedded in early-stage combat patterns were identified during model development (Section 3), where motif bigrams consistently improved forecast performance across scenarios. The layered design supports decision-making by linking elevated casualty scores to specific motif patterns, thereby allowing risk signals to be traced to observable tactical behaviours. This structure offers both analytical depth and operational transparency.

4.2. User Interface and Component Roles

SCOPE provides an interactive user interface designed to integrate predictive outputs, interpretive plots, and diagnostic tools (Figure 2). The interface mirrors the layered structure of the modelling framework and provides analysts with a transparent and responsive environment for tracking evolving mission risk. Insights generated through this interface can then be communicated to decision-makers in real time, supporting earlier and more informed operational responses.

To support explanation of the early warning system, the live dashboard has been reformatted into a static, multi-panel layout suitable for printed presentation. Figure 2 reflects a typical analyst view of the system, with each panel corresponding to a core function of the SCOPE framework. The central panel displays the predicted *Casualty Score* from the pooled MARS model. This score updates dynamically as each new time window is processed, allowing analysts to monitor changes in risk levels over time. The panel adjacent to the prediction presents interpretive plots derived from the GAM model, highlighting how specific motifs influence the predicted Casualty Score and supporting clearer understanding of underlying risk drivers. A third panel in Figure 2 presents a variable importance summary based on the RF model. This ranking of predictors by influence serves as a practical audit of model reasoning and highlights which motifs most affect predicted outcomes. Each component updates automatically in response to new event data.

A key feature of the interface is the manual input calculator, which allows analysts to enter observed on-field losses, such as the number of tanks, IFVs, and infantry, to compute a value of the *Casualty Score*. This enables real-time comparison between observed and predicted outcomes, helping decision-makers assess whether the mission is tracking better or worse than expected. The system flags the result using colour-coded feedback (green, orange, red), providing an immediate and intuitive diagnostic summary. Input ranges are constrained by operational limits set in the background to prevent implausible entries.

Decision-makers can filter results by scenario and time window, supporting both exploratory analysis and operational use in high-tempo environments. The manual input calculator provides an estimate that offers a point of comparison against model predictions, supporting rapid assessment of whether the unfolding mission trajectory aligns with anticipated risk or signals an elevated threat. Together, the SCOPE interface and modelling layers form a cohesive early warning system by translating local event series into interpretable risk signals and enabling timely evaluation of mission status.

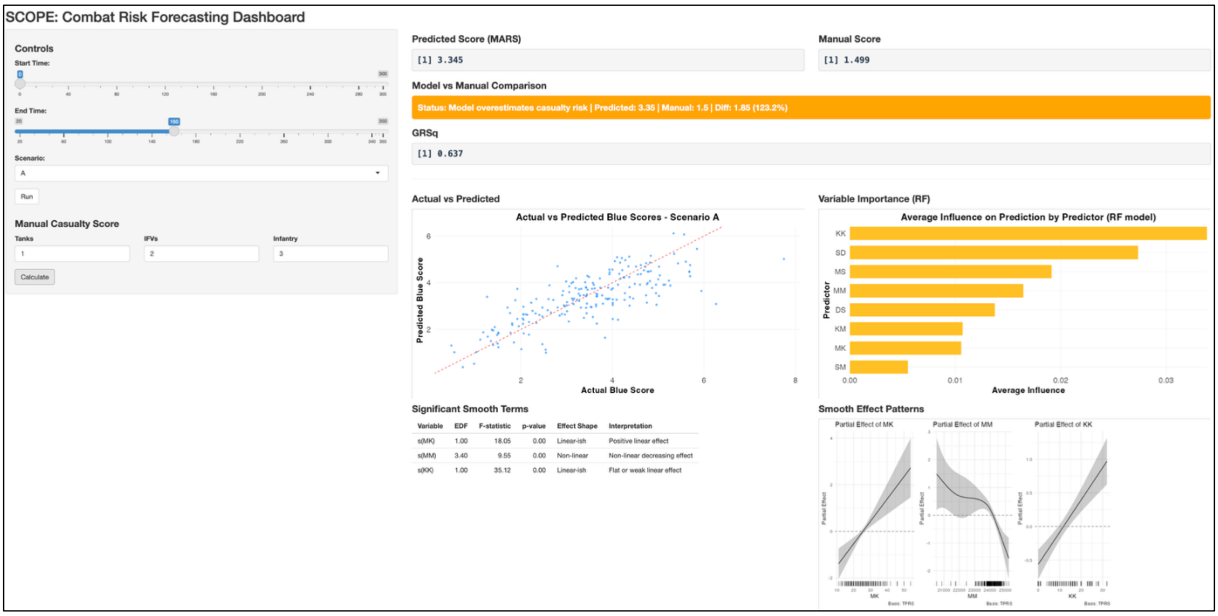

Figure 2. Summary panel of the SCOPE interface, displaying predicted risk scores, interpretive plots, and variable importance.

5. CONCLUSION

This study introduced an integrated early warning system for combat simulations that combines predictive accuracy with interpretive clarity and diagnostic transparency. By extracting motif bigrams from tactical event series and linking them to casualty outcomes via the *Casualty Score*, SCOPE identifies short-term patterns that signal elevated operational risk. The modelling framework layers a pooled MARS model for outcome prediction with Generalised Additive Models (GAMs) for interpretive structure, enabling robust forecasts while revealing the specific predictors contributing to shifting risk during combat progression.

SCOPE operates incrementally, updating predictions across successive time windows to reflect evolving mission conditions. The interactive interface unifies predictive outputs, interpretive plots, and manual scoring tools. Analysts can track risk trajectories, compare predictions against observed field losses, and audit the influence of individual event motifs in real-time. As shown in Section 3.3, SCOPE consistently delivered reliable performance across multiple scenarios, with predictive signal extraction generalising well while remaining sensitive to scenario-specific variation, to support robustness and operational relevance.

The next step is to assess system performance in more realistic and under time-sensitive conditions. Current findings establish the framework as stable and practically valuable within simulation environments. Future in-situ testing will allow evaluation of usability, responsiveness, and tactical utility as risk patterns unfold under operational constraints. Embedding the system within live or near-real-time simulations will also support refinement of dynamic thresholds, alert mechanisms, and feedback loops, ensuring alignment with the demands of high-tempo missions. With the foundational framework now demonstrated, SCOPE is positioned for deployment into applied testing environments. Such a transition will provide the empirical grounding necessary to move from proof-of-concept to eventual operational readiness.

ACKNOWLEDGMENTS

The Commonwealth of Australia (represented by the Defence Science and Technology Group) supported this research through a Defence Science Partnership agreement. We would like to thank Dr Maria John and Dr Lisa Blumson for their invaluable input into this research through expert insight.

REFERENCES

Brown, M. E., Mugo, S., Petersen, S., & Klauser, D. (2022). Designing a Pest and Disease Outbreak Warning System for Farmers, Agronomists and Agricultural Input Distributors in East Africa. Insects (Basel, Switzerland), 13(3), 232.

Chau, W., Gill, A., & Grieger, D. (2017). Using combat simulation and sensitivity analysis to support evaluation of land combat vehicle configurations. In *22nd International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand* (pp. 536-42).

Chmielewski, M., Kukiełka, M., Pieczonka, P., & Gutowski, T. (2020). Methods and analytical tools for assessing tactical situation in military operations using potential approach and sensor data fusion. *Procedia Manufacturing*, *44*, 559-566.

Gill, A., Grieger, D., Wong, M., & Chau, W. (2018, December). Combat simulation analytics: Regression analysis, multiple comparisons and ranking sensitivity. In *2018 Winter Sim. Conference (WSC)* (pp. 3789-3800). IEEE.

Grieger, D., Wong, M., Tamassia, M., Torres, L., Giardina, A., Vasa, R., & Mouzakis, K. (2021). Towards the identification and visualization of causal events to support the analysis of closed-loop combat simulations. In Data and Decision Sciences in Action 2: Proceedings of the ASOR/DORS Conference 2018 (pp. 295-312). Springer International Publishing.

Grünwald, P. D. (2007). The minimum description length principle. MIT press.

Hotz, I., Hanisch, A., & Schulze, T. (2006, December). Simulation-based early warning systems as a practical approach for the automotive industry. In Proceedings of the 2006 Winter Simulation Conference (pp. 1962-1970). IEEE.

Lin, J., Keogh, E., Lonardi, S., Patel, P., 2002. Finding motifs in time series. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2nd Workshop on Temporal Data Mining, pp. 53–68.

Lyu, X., Liu, C., Xie, Z., & Xu, C. (2023, October). An Analysis Method of Target Value Considering Combat Phases. In China Conference on Command and Control (pp. 75-88). Singapore: Springer Nature Singapore.

Masterson, M. J. (2004). Using assessment to achieve predictive battlespace awareness. Air & Space Power Journal [Chronicles].

Mathais, Q., De Malleray, H., Nguyen, C., Weghel, L. L., Boussen, S., & Bordes, J. (2025). Enhancing combat casualty care in military medicine: the potential of early warning systems and wearable biosensors in large-scale warfare. *BMJ Mil Health*.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. Nature (London), 461(7260), 53–59.

Schubert, J., Moradi, F., Asadi, H., Luotsinen, L., Sjöberg, E., Hörling, P., Linderhed, A., & Oskarsson, D. (2015). Simulation-based decision support for evaluating operational plans. Op. Res. Perspectives, 2, 36–56.

Tiong, J., Bogomolov, T., & Chiera, B. (2023). Markov chain modelling of causality in anomalous and non-anomalous combat event series. In *25th International Congress on Modelling and Simulation* (pp. 953–959). Modelling and Simulation Society of Australia and New Zealand.

