Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 29, 2021
A Regional Conference of the Society for Industrial and Systems Engineering

Data Analytics Development from Military Operational Data

James Downey[1], Zachary Ellis[1], Ethan Nguyen[1], Charlotte Spencer[2], and Paul Evangelista[1]

[1]United States Military Academy
Department of Systems Engineering
West Point, NY

[2]United States Military Academy
Department of Mathematical Sciences
West Point, NY

Corresponding author's Email: Zachary.Ellis2@westpoint.edu

Author Note: Cadets Downey, Ellis, Nguyen, and Spencer are all seniors at the United States Military Academy (USMA) participating in a year-long capstone project under the supervision of Colonel Paul Evangelista.

Abstract: Each year, the National Training Center (NTC) located at Fort Irwin, California, hosts multiple Brigade-level rotational units to conduct training exercises. NTC's Instrumentation Systems (NTC-IS) digitally capture and store characteristics of movement and maneuver, use of fires, and other tactical operations in a vast database. The Army's Engineer Research and Development Center (ERDC) recently partnered with Training and Doctrine Command (TRADOC) to make some of the data available for introductory analysis within a relational database. While this data has the potential to expose capability gaps, uncover the truth behind doctrinal assumptions, and create a sophisticated feedback platform for Army leaders at all levels, it is largely unexplored and underutilized. The purpose of this project is to demonstrate the value of this data by developing a prototype information system that supports post-rotation analytics, playback capabilities, and repeatable workflows that measure and expose ground-truth operational and logistical behavior and performance during a rotation. The Army modeling and analysis community will use these products to systematically curate and archive the database and enable future analysis of the NTC-IS data.

Keywords: National Training Center (NTC), Data Analytics

1. Introduction and Related Work

Businesses and organizations around the globe are constantly looking to find ways to improve their performance and gain a comparative advantage over the competition. With the rise in technological capabilities and digital documentation, many have turned to a data-driven approach, utilizing massive databases to explore patterns of success, uncover performance trends, and help real-time decision making (Ofoghi, 2013). Professional sports present a perfect example as they routinely devote billions of dollars to data collection, storage, analysis, and visualization (Ricky, 2019). Despite the proven effectiveness of these techniques, the United States Army has fallen behind in its own data exploration and data usage capabilities.

The National Training Center located at Fort Irwin, California, serves as the Army's preeminent training environment for Brigade-level operations; consequently, it provides a unique opportunity to collect vast amounts of data. This data collected at NTC could expose capability gaps within our formations, uncover the truth behind doctrinal assumptions, and create a sophisticated feedback platform for Army leaders at all levels; however, the data is largely unexplored and underutilized. Researchers and data analysts are looking to the NTC database to provide instantaneous feedback systems, performance trends, and probabilistic statistics to Army leaders but require more robust analytical workflows to do so.

Several key stakeholders throughout the Army and third-party companies recognize the need to modernize the data collection and analysis derived from NTC rotational data. The first two primary stakeholders include the Army's Engineer Research and Design Center (ERDC) and U.S. Army Training and Doctrine Command (TRADOC). Both are interested in designing systems capable of archiving and analyzing the NTC data. The two compiled all historical rotational data into a relational database for introductory analysis while designs for a more permanent data management system are developed. The United States Military Academy at West Point was granted access to 30 rotations worth of data in the database, five of which were used to develop a prototype information system and build a proof of concept for ERDC, TRADOC, and other stakeholders interested in further analyzing the data.

The third group of stakeholders in this project are the users. Potential users include operational leaders, the acquisitions community, simulators and modelers, and doctrine writers. Each of these communities could greatly benefit from basic statistics on Soldiers, key weapon systems, vehicles, etc. as well as correlating factors with tactical and operational success. Data users should also be able to use a data-driven approach to inform their decisions, validate or update their assumptions, and draw unintuitive conclusions. Our team's workflows and data analytics have proven that the NTC data has the potential to meet all of these capabilities and more.

Literature related to this research spans several domains. The past decade has led to many technological advances enabling companies and organizations to collect, measure, and analyze large volumes of data (Passfield and Hopker, 2017). Significant resources are allocated to data mining, regression analysis, and other data analysis techniques to develop performance predictions, identify trends, and correlate events to build more effective and competitive business models in many sectors (Haas and Mortenson, 2016). Professional sports, a major consumer and user of data analytics, is expected to spend upwards of $4 billion by 2022 in the collection, storage, and analysis of data (Ricky, 2019). Sports analysts use a variety of techniques to identify patterns of success and performance trends; these analytics commonly inform coaches, players, general managers, and others in their pre-game and game-time decisions (Ofoghi, 2013).

Within the past twenty years, the National Training Center has increased its data collection capabilities through the implementation of the CTC-IS system (U.S. Army NTC, 2020a). NTC's simulation center currently records data on entity locations, shot pairings, key events, order of battle, and battle damage assessments (U.S. Army NTC, 2020a). This information is used to provide anecdotal evidence within after action reviews (AARs) as units complete their training rotations at NTC (U.S. Army, 2020a). While these AAR products are useful, research suggests that the current analytical products available in AARs are lacking in their ability to uncover hidden trends, metrics, and other key lessons learned at the Company and Battalion level during a rotation (Schoellhorn, 2020 and U.S. Army NTC, 2020a).

In partnership with RAND, a federally funded research and development center, Andrew Cady explored how data from the NTC can be used to support the acquisitions community (2017). Cady derived measures of effectiveness for the probability of hits, rates of fire, unit dispersion, and unit speed from the database which he believed could help the acquisitions community field new technologies to fill in the capability gaps exposed from data analysis (2017). Dana Goulette also discussed the capabilities of NTC data to systematically measure and assess unit performance at the training center in his Naval Postgraduate School thesis (1997). Goulette's assessment model utilizes the relational database to conduct post-rotation analysis, trend identification, and compare unit performance (1997). Despite Goulette's early work in creating standardized measures of performance, these metrics have largely gone unexplored and training technologies have since vastly improved.

Authors of a 2020 White Paper from the Operations Group at NTC stated significant portions of NTC data are withheld because of "reporting restrictions" (U.S. Army NTC, 2020b). The report says, "because Brigade Combat Teams (BCTs) pay-to-play, we are restricted from publishing specific information that could embarrass leaders or restrict experimentation" (U.S. Army NTC, 2020b). The team notes that many of the mistakes, failures, and lessons learned occur repeatedly at NTC, but until Army leadership is ready to receive candid feedback about unit performance in pre-deployment training, the Army cannot reach its full potential (U.S. Army NTC, 2020b).

In a similar White Paper, authors explore the potential for correlating data across rotations to build out trends, patterns, and other metrics exposed by the NTC-IS data (U.S. Army NTC, 2020a). The authors claim that "we should be able to correlate data with fires, observers, accuracy, and lethality. All of the data is there, but we are lacking the tools, talents, and military guidance to develop the data and analytics to take us further and make us better" (U.S. Army NTC, 2020a). This paper describes the methodology, workflows, and resulting data analytics developed by the West Point team in order to provide a framework of analysis that future analysts can utilize to further explore and process the data.

2. Methodology and Workflows

Essential to the success of this project is the development of analytic workflows that effectively derive descriptive combat metrics that can improve small unit performance. This project presents several risks that are important to mitigate, due to the breadth of information available in the relational database and the number of stakeholders with vested interests. This project is both conceptually and technically risky due to its manipulation of data and the development of new systems that aim to improve training and Soldier effectiveness. Instrumentation systems at the NTC provide the data used to create repeatable workflows; as a result, any change to these instrumentation systems would alter the input of already created workflows. To mitigate this risk, the workflows will remain as dynamic and flexible as possible. Given the five available rotations of data, scripts will be tested to ensure flexibility and minimize project-specific risk.

This Capstone project assumes technical risk since data is being converted and transferred across different operators. Many of the technical risks include losing access to the NTC data, not having the coding expertise to appropriately analyze the data, and potentially losing any data, workflows, metrics, and analytics that are developed. Creating technical redundancies by

exporting all relevant files to cloud-based servers, version control, and practicing good coding and documentation practices can help mitigate many of these technical risks.

The primary purpose of this project is to develop a prototype information system that produces well-documented and repeatable workflows. The relational database is sophisticated and requires extensive architectural mapping, and the prototype information system must sort, analyze, and visualize the data within. The repeatability of developed workflows is crucial to enabling follow-on analysis by outside agencies. The functional decomposition for this system is shown in Table 1. The lowest level functions represent the workflows designed for this system.

Table 1. Functional Decomposition

Fundamental Objective: Uncover Hidden Trends, Patterns, and Metrics Exposed in the NTC Data		
Objective 1.0 Create Usable Data	Objective 2.0 Develop Repeatable Rotational Analytics	Objective 3.0 Develop Repeatable Multi-Rotational Analytics
1.1 Document database metadata 1.2 Identity tables and views with high-value data 1.3 Query and export data	2.1 Determine order of battle 2.2 Calculate rotation-specific direct fire lethality* 2.3 Calculate rotation-specific indirect fire lethality* 2.4 Provide playback ability 2.5 Develop visual aids for analytics*	3.1 Determine quantities and types of vehicles per rotation 3.2 Calculate multi-rotation direct fire lethality* 3.3 Calculate multi-rotation indirect fire lethality* 3.4 Calculate performance metrics based on positional data* 3.5 Develop visual aids for analytics*

* = documented workflow outlined in this paper; all other functions have prototype workflows but are not specifically discussed within this paper.

Figure 1 outlines the current process used to transform raw NTC data into finished analytical products. To build context, unstructured data from NTC archives are reviewed to build situational awareness for each rotation. This unstructured data includes AARs, timelines, and key events. The data servers provide the contextual information required to understand events in a particular rotation and highlight which analytics should be created to further develop the database. From the relational database, an initial query was created to list the meta-data for all tables and views **(Function 1.1 Document database Metadata and Function 2.2 Identify tables with high-value data)**. Following this initial query, SQL scripts are used to query high-value data tables from the relational database and export them to flat files **(Function 1.3 Query and Export data).** The bulk of analysis is then conducted with Cygwin, R, and Google Earth to create analytical workflows. Cygwin, a Linux emulator for Microsoft Windows, is primarily used to filter, sort, and parse the data. R is primarily used to further filter the data, perform statistical analysis, and create visual aids to display our analytical models. Google Earth is primarily used as background software to link positional data with terrain data to display a prototype playback capability. As a result of these processes, the team has developed several queries and analytic scripts in these applications that fetch high-value data and create analytical products that incorporate single and multiple rotations (**Objectives 2.0 Develop Repeatable Rotational Analytics)**.

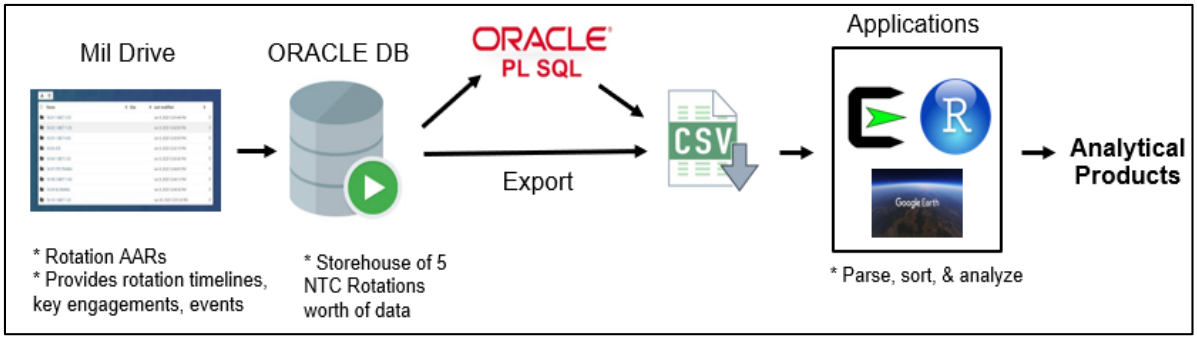

Figure 1. Project Workflow

A brief discussion of several key workflows follows. These selected workflows highlight some of the most salient analytical methods designed to explore this data. (**Functions 2.2 Calculate Direct Fire lethality)** The direct fire workflow

calculates the volumes of fires and specifies the Damage Battle Assessment (BDA) of each engagement. This workflow summarizes shooting events between opposing forces (OPFOR) and blue forces (BLUFOR, the rotational training unit), developing visualizations for mass of fires and lethality. To achieve this function, SQL queries isolate information specific to a single rotation and exports that information to flat files so that several rotations can be analyzed comparatively. A summation of the number of rounds fired is used as a measure of the volume of direct fire, while recordings of shot pairing hits and damage assessments are used as proxies for accuracy and lethality. In this workflow, lethality is defined as the total number of damage assessments "kills" over the total volume of fire.

A similar workflow determines the haversine distance between all shooters and their respective targets. The workflow queries all direct fire shot pairings within in the database and exports them into flat files. Once loaded into R, various functions calculate the exact distances between shooters and their respective targets which are then plotted in a series of box plots.

(Function 3.3 Calculate Indirect Fire lethality) The indirect fire (IDF) workflow develops two visualizations to demonstrate the use of indirect fires across several rotations. Within relational, a series of database views with IDF data are spooled into flat files using an SQL query. The IDF views are converted into flat-files with important location, time, and identifying data. These flat files are loaded into R where a series of data cleaning and transformation functions filter, organize and calculate statistics and tables for analysis and visualization. The cleaned and calculated statistics support the creation of plots, leaflet maps, and eventually a Shiny App that combines the visualizations. The visualizations are comparisons to show BLUFOR Commanders the effective or ineffective use of fires throughout a rotation compared to the OPFOR.

(Function 3.4 Calculate performance metrics using positional data) The positional data workflow uses locational data to determine the speeds and distances of all vehicles in a given rotation. A SQL Query spools a complete flat file that contains the positional changes of thousands of NTC vehicles based on a 'UTC' time entry – this is possible because each vehicle holds a specific entity identifier. A distance haversine function is then applied—this function uses the longitude, latitude, and radius of the earth to determine how far a vehicle has traveled between time entries. This workflow provides units with information about specific information for different vehicle and unit types.

3. Results and Discussion

Select plots reinforce the analytical potential of the information system. The development of these plots provides a 'proof of principle' of battlefield analytics that can be used to provide feedback to future units. The boxplots in Figure 2 depict direct fire data recorded across three different NTC rotations **(Function 3.2 Calculate multi-domain direct fire lethality).**

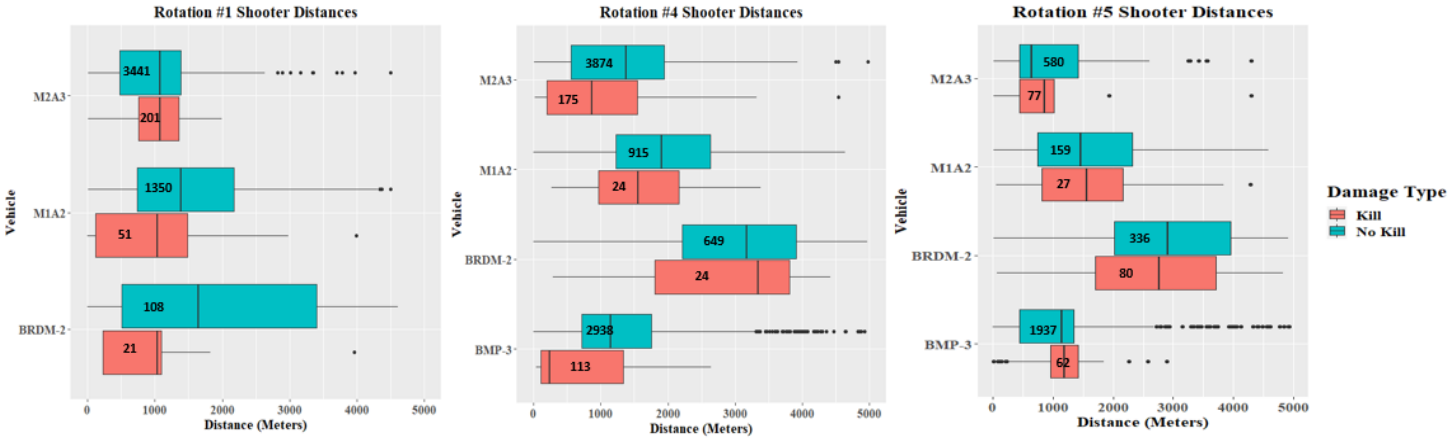

Figure 2. Direct Fires Boxplot

The boxplot shows the distribution of shot pairs with distances between shooters and their respective targets. On the y-axis, different vehicles are compared (M2A3 Bradley infantry fighting vehicle, M1A1 Abrams main battle tank, BRDM OPFOR fighting vehicle, etc). Conversely, on the x-axis, exact distances between these vehicles and their targets are displayed. The blue bar represents the distribution of shots that did not result in kills, while the red bar below it represents

successful kills. This provides valuable insight to ground commanders about the lethality of vehicles and highlights effective ranges that vehicles should be firing from in order to be successful. With a brief script in R, the same visualization was created across three rotations. The repeatable nature of the script verifies that the workflow is scalable for future analysis.

The graphs in Figure 3 display the lethality and volume of fires workflow developed for a singular rotation **(Function 2.5 Develop visual aids for analytics)**. In this example, the volume of fires is depicted on the left for the most prominent vehicles during Rotation 5 at NTC. The graph on the right depicts shots that landed on a target and identifies specific shots which resulted in a change of damage assessment. Based on visual analysis, it is clear that during this rotation, accuracy for both the OPFOR and the BLUFOR required improvement. A small proportion of shots taken landed on target, and an even smaller proportion proved to be lethal. This type of graphical depiction can be replicated for multiple rotations in order to identify potential trends across different rotational units.

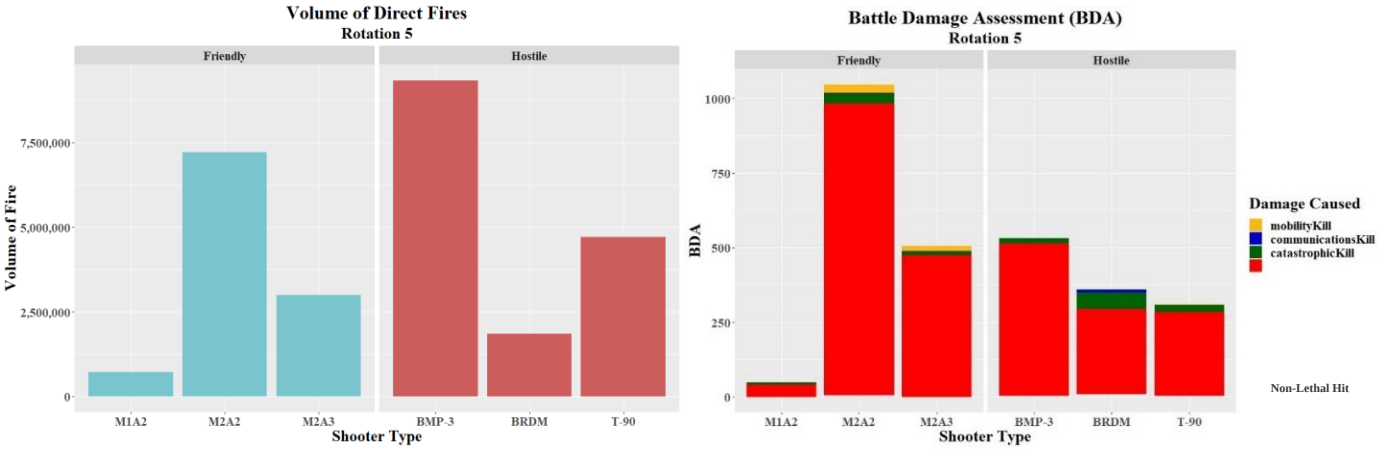

Figure 3. Volumes and Lethality of Direct Fires

Analysis of contextual data and historical unit reports from the rotation reveals that the volume of indirect fires from the opposing forces consistently outmatches the rotational units **(Function 3.3 Calculate Indirect fire lethality)** The graphs on the right in Figure 4 demonstrate BLUFOR's lack of indirect fire volume. The OPFOR's ability to better utilize the IDF demonstrates their ability to plan and execute missions at all echelons both before and during combat engagements. In the map on the left, the discrepancy in the volume of fires is again demonstrated. However, through spatial analysis, it is noted that the IDF placement for OPFOR is much more concentrated and centered around key avenues of approach.

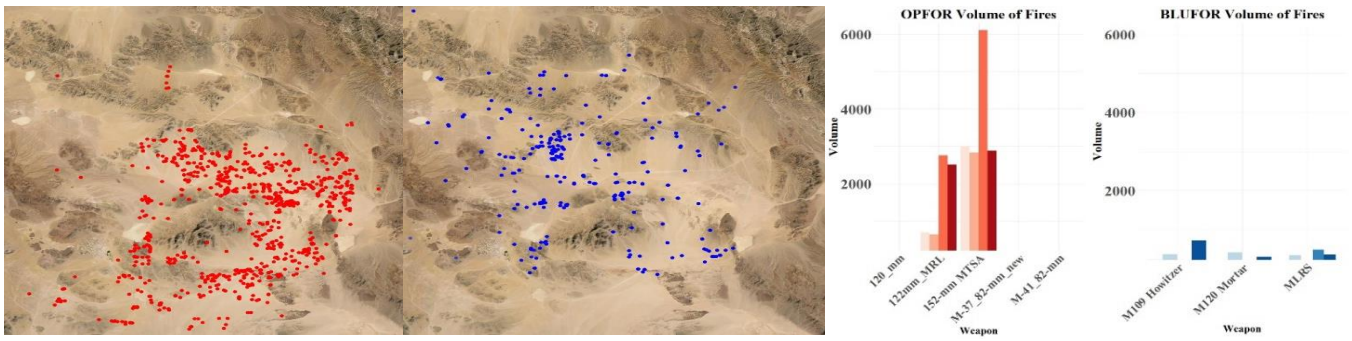

Figure 4. Volume of Indirect Fire

6. Conclusion and Future Work

Data analytics at the National Training Center are far behind the professional sports world and other data-driven communities in terms of its data exploration, analysis, and visualization capabilities. Analysis of data from training and contingency operations has the potential to expose capability gaps, uncover the truth behind our doctrinal assumptions, and

create a sophisticated feedback platform for our Army leaders to use at every level. This project showcases the value of exploring this data by establishing prototype analytics, rudimentary playback capabilities, and other repeatable workflows which the Army's modeling and analysis community can use as a framework for future and deeper analysis.

Many organizations and researchers have claimed interest to further consolidate, explore and analyze the NTC data. The Army's Engineer Research and Development Center (ERDC) is interested in migrating the NTC data into a new, consolidated database repository to perform many of these functions. ERDC's goal is to provide an easy-to-access computational space for Army researchers, simulators, and data scientists to further clean, mine, and analyze the data.

Future data-driven analysis does face limitations. All workflows developed thus far are specifically tailored to input provided by intstrumentation systems at NTC. Other CTCs and training entities will need to develop similar data collection processes in order to benefit from the analytical workflows developed by the current relational database.

An entire host of customers stand to potentially benefit from these continuations of our work. Entities within training, evaluation, acquisition, analysis, simulation, intelligence, and academic communities all stand to benefit from the captured NTC data and its descriptive analytics. The operational community wishes to explore hidden capability gaps within the Army's formations, challenge pre-existing assumptions, and re-evaluate Army doctrine based on a data-driven approach from this data repository. The acquisitions community is primarily concerned with the potential capability gaps exposed by the NTC data and then outfitting the Army with new or improved technology to fill those gaps.

The consolidation and analysis of the NTC data has the potential to encourage analysts to continue exploration across different environments and training locations. Massive amounts of data are also collected at the other combat training centers (CTCs) as well as on deployments. Analysis methods presented in this paper could leverage this data and provide Army leaders more expansive insights across a broad spectrum of training and contingency operations. Regardless of the user, the data captured at these CTCs and during other operations have the potential to explore hidden capability gaps, challenge our pre-existing assumptions about warfare, and uncover the truth behind our operational and logistical competencies making the Army a more lethal, survivable, and effective fighting force.

7. References

Cady, A. (2017). *Using the National Training Center instrumentation system to aid simulation-based acquisition*. RAND.

Goulette, D. (1997). *Training assessment and modeling subjective data encapsulation for the National Training Center*. Monterey, California. Naval Postgraduate School.

Haas Martine & Mortensen Mark (2016). Leading Teams: The Secrets of Great Teamwork. Harvard Business Review https://hbr.org/2016/06/the-secrets-of-great-teamwork

Ofoghi, B., Zeleznikow, J., MacMahon, C., & Raab, M. (2013). Data Mining in Elite Sports: A Review and a Framework. *Measurement in Physical Education and Exercise Science*, 17(3), 171–186. https://doi.org/10.1080/1091367x.2013.805137

Passfield, L., & Hopker, J. (2017). A Mine of Information: Can Sports Analytics Provide Wisdom From Your Data? *International Journal of Sports Physiology and Performance*, 12(7), 851–855. https://doi.org/10.1123/ijspp.2016-0644

Ricky, A. (2019, January 31). Council post: How data analysis in sports is changing the game. Retrieved March 09, 2021, from https://www.forbes.com/sites/forbestechcouncil/2019/01/31/how-data-analysis-in-sports-is-changing-the-game/?sh=2ed8eef13f7b

Schoellhorn, B.P. (2020). Preventing the Collapse: Fighting Friction after First Contact at the National Training Center. *Military Review*, 100(2), p. 6.

U.S. Army National Training Center. (2020a). *Combat Training Center-Instrumentation System (CTC-IS) Data & Analytics* [White paper]. U.S. Army FORSCOM.

U.S. Army National Training Center. (2020b). *Utilizing the Rotational Training Units (RTUs) Data* [White paper]. U.S. Army FORSCOM.