

Attributing Mission Performance to Submarine Subsystems

Jonathan E. Proule

Undersea Warfare Mission Engineering and Analysis Department

3 January 2024



**Naval Undersea Warfare Center Division
Newport, Rhode Island**

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

ADMINISTRATIVE INFORMATION

This technical memo was prepared under NUWC Division Newport Network Activity No. 100001936442 0010, "FY23 C60 SSN(X) Dashboard," principal investigator Jonathan E. Proule (Code 60A). The sponsoring activity is the NUWC Division Newport Internal Investment Program managed by Derek K. Potvin (Code 00X) of the NUWC Division Newport Strategic Planning Office.

ACKNOWLEDGMENTS

The author thanks Martin J. Guyotte (Code 60A) and Christopher T. Kona (Code 60E) for their support through the early stages and continued development of the Mission Effectiveness (ME) Dashboard tool described herein. Additionally, the author expresses gratitude to PMS 4XX for financial support and guidance helping shape the processes described in this technical memo.

REPORT DOCUMENTATION PAGE

1. REPORT DATE 03-01-2024		2. REPORT TYPE Technical Memorandum		3. DATES COVERED	
				START DATE 30-06-2019	END DATE 30-06-2023
4. TITLE AND SUBTITLE Attributing Mission Performance to Submarine Subsystems					
5a. CONTRACT NUMBER N/A		5b. GRANT NUMBER N/A		5c. PROGRAM ELEMENT NUMBER N/A	
5d. PROJECT NUMBER N/A		5e. TASK NUMBER N/A		5f. WORK UNIT NUMBER N/A	
6. AUTHOR Jonathan E. Proule					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Undersea Warfare Center Division 1176 Howell Street Newport, RI 02841-1708				8. PERFORMING ORGANIZATION REPORT NUMBER TM 19-137	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Sea Systems Command PEO SSN, Simeon Rosenthal Building 201 614 Sicard Street Southeast Washington Navy Yard, Washington, DC			10. SPONSOR/MONITOR'S ACRONYM(S) NAVSEA PEO-SSN		11. SPONSOR/MONITOR'S REPORT NUMBER(S) N/A
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES N/A					
14. ABSTRACT By employing statistical emulators in place of higher-fidelity models, the Mission Engineering & Analysis Department of the Naval Undersea Warfare Center Division Newport has been able to quickly assess the military utility of different submarine components. This process supports the efforts of naval architects as they develop different conceptual submarine designs and identify which components to include as part of those designs.					
15. SUBJECT TERMS Surrogate Modeling, Meta-Model, Emulators, Model-Agnostic Interpretability, Counter-Factuals, Data Visualization					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 15	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			
19a. NAME OF RESPONSIBLE PERSON Jonathan E. Proule				19b. PHONE NUMBER (Include area code) 401-832-6472	

TABLE OF CONTENTS

Section	Page
1. BACKGROUND AND MOTIVATION.....	1
2. INTRODUCTION.....	2
3. MISSION EFFECTIVENESS DASHBOARD.....	3
3.1 Prediction.....	3
3.2 Inference and Visualization.....	4
3.3 Operational Context.....	6
3.4 Application and Mapping Models.....	7
4. CONCLUSION.....	10

LIST OF FIGURES

Figure	Page
1 The ME Dashboard landing page.....	2
2 Data upload window screenshot for a single mission in the ME Dashboard.....	4
3 Outlier investigation window screenshot of the ME Dashboard.....	6
4 Screenshot of the ME Dashboard landing page.....	7
5 Submarine concept evaluation flowchart.....	8
6 Information flow from architect to analyst.....	9

LIST OF ABBREVIATIONS AND ACRONYMS

ME	Mission Effectiveness
NUWC	Naval Undersea Warfare Center
SHAP	SHapley Additive exPlanations

1. BACKGROUND AND MOTIVATION

Submarine design is a multifaceted process requiring extensive effort spanning across various disciplines and organizations. Part of this effort is deciding which missions a submarine is expected to conduct and assessing its efficacy, or mission effectiveness, in those missions via constructive simulation. This approach—characterizing a notional platform and simulating its performance in stressing tactical situations—has been and continues to be the foundation of most analytic products produced by the warfare analysts of the Undersea Warfare Mission Engineering and Analysis Department of the Naval Undersea Warfare Center (NUWC) Division Newport. It allows analysts to quantify the effect of different aspects of the platform, such as its acoustic profile or maneuverability, in terms of mission effectiveness. They can then answer questions such as “What sort of payloads should the submarine have to meet its mission goals?” or “What sensing capability is required to meet a threshold of mission effectiveness?”

To ensure that a submarine concept can meet mission requirements, the conceptual submarine must first be adequately characterized, typically by defining features such as the type of sensor systems it possesses, how it maneuvers, its acoustic profile, and its weapon composition. It can then be assessed across relevant, stressing scenarios, many with competing objectives. For instance, a conceptual submarine may favor the use of torpedoes in one mission while preferring the use of missiles for another. Stowage aboard the platform is limited due to space, so the decision must be made as to how to balance both the number and type of weapon to meet competing mission requirements. By simulating submarine performance across various missions, analysts can quantify the tradeoffs associated with each design decision.

While much is to be gained from this approach, stakeholders may find themselves limited when attempting to gauge the mission effectiveness of a platform at input levels that were not originally run as part of the constructive simulation. As an example, if three levels of submarine speeds (5 knots, 15 knots, and 25 knots) were simulated as part of a study and the customer was interested in what the associated mission effectiveness at 10 knots would be, the customer is relegated to mental means of interpolation to estimate the quantity. While this is achievable in lower dimensional space, it quickly grows difficult in the presence of additional attributes (such as submarine acoustic profile, maneuverability, and sensing capability), as does the visualization of these estimates. Furthermore, that customer may not have the operational context and assumptions (threat capability or environmental conditions) readily available to make informed design decisions.

It is through the department’s Mission Effectiveness (ME) Dashboard that analysts and customers can rectify these shortcomings and facilitate the concept assessment process. This technical memo will review the capabilities of the ME Dashboard, the mapping models that translate architectural design components into ME Dashboard inputs, and its application in the evaluation of a notional submarine concept with the goal of attributing mission performance to the submarine concept’s subsystems.

aspects of the ME Dashboard, we will return to this scenario to review mapping models and see how the ME Dashboard can be employed to address the idea of attributing mission performance to submarine subsystems.

3. MISSION EFFECTIVENESS DASHBOARD

3.1 PREDICTION

The foundational competency of the ME Dashboard is its predictive capabilities. Derived from the ideas of response surface methods associated with classical experimental design techniques, the ME Dashboard houses a collection of statistical and machine learning models that act as fast-running surrogates to the slower-running, higher-fidelity models. These higher-fidelity models are used to simulate submarine performance for each one of the missions it is expected to conduct. The inputs and outputs from these simulations are captured as the covariates and response variable used in the training/fitting of the surrogate models in a supervised fashion.

These surrogates provide estimates of what the higher fidelity models would produce if they were used to simulate over inputs of interest. The ME Dashboard houses the following types of surrogate models:

- Elastic net regularized regression
- Generalized additive models
- Binary logistic regression
- Beta regression
- Binary Gaussian processes
- Treed Gaussian processes
- Gradient-boosted trees
- Multiadaptive regression splines
- Feed-forward neural networks

Which surrogate model is applied to which mission depends on the type of input data that is under consideration. Lower-dimensional, all-numeric data tends to be represented well by Gaussian processes; this class of surrogate models can capture the predictive variance of numeric data that does not exhibit any step function-like behavior. Conversely, data with many categorical features may be better fit with tree-based models. These surrogate models may be stacked or combined/extended via boosting, bagging (bootstrap aggregation), or other ensemble methods and may require additional feature engineering on the original data produced from the simulation run with the higher fidelity model. The only constraints on the surrogates is that they are relatively quick to load and operate (does not exceed five seconds) and they generate predictions with low (<3%) generalizable error (a measure of how accurately an algorithm can predict outcome values for previously unseen data). This generalizable error is assessed via different kinds of nested cross-validation or other forms of out-of-sample assessments such as train-test-validation sets and bootstrapping. To help reduce this generalizable error and foster the creation of highly predictive surrogate models, the higher-fidelity models typically simulate over specialized experimental designs that allow nonlinear effects to be realized. These experimental

designs are often space-filling designs that offer exceptionally strong coverage over the tradespace of interest, or they may be a hybrid between space-filling and classical experimental designs (fractional factorial or D-optimal designs).

Through its user interface, the ME Dashboard offers users the ability to upload data, or manually enter it within the browser, and then generate predictions with the click of a few buttons (Figure 2). There are no command-line interactions or coding requirements, and users can download predictions for further analysis.

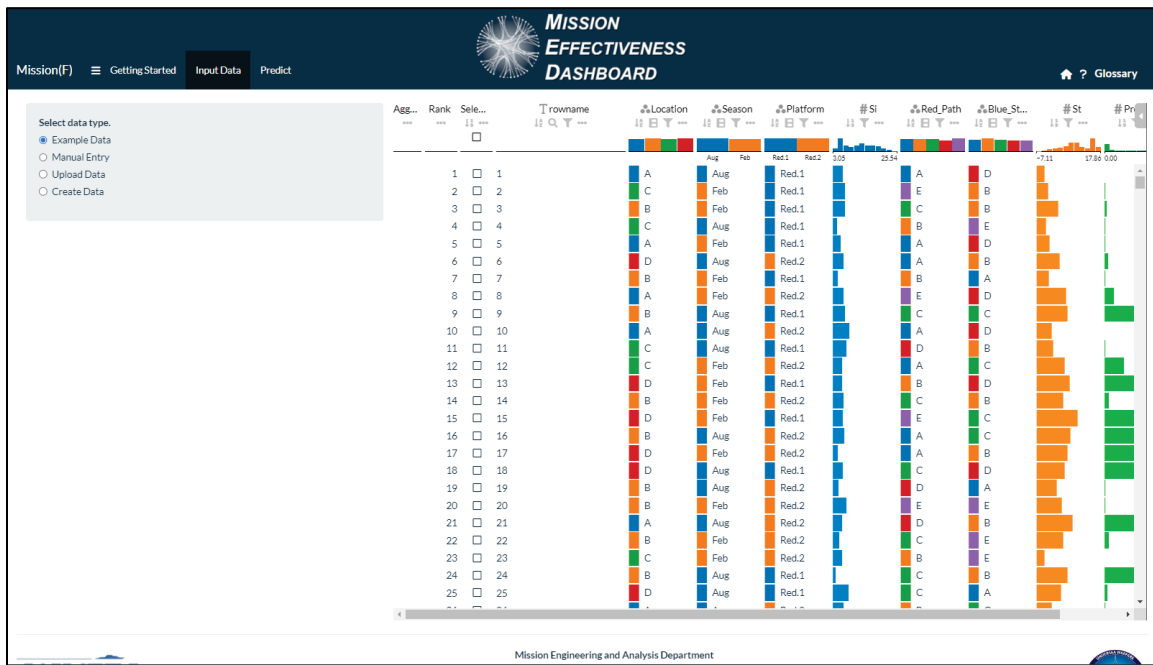


Figure 2. Data upload window screenshot for a single mission in the ME Dashboard.

3.2 INFERENCE AND VISUALIZATION

In addition to downloading the predictions generated from the surrogate models within the ME Dashboard, users can investigate the predictions via various visualizations and inference methods. They can employ any of the following within the ME Dashboard to graphically depict their predictions or the example datasets that come with each housed mission:

- Heatmaps
- Treemaps
- Scatterplots and scatterplot matrices
- Sunburst plots
- Parallel coordinate plots
- Correlation matrices/plots
- Histograms
- Density plots
- Boxplots

While these techniques are a useful means to quickly identify and plot trends and gradients within the data, users are also supplied with options to quantify these trends via more rigorous means like the following:

- Individual conditional expectation plots
- Partial dependency plots
- Beeswarm and waterfall plots to visualize SHapley Additive exPlanations (SHAP)
- Breakdown plots
- Variable importance plots
- Sobol indices plots
- Friedman's H-statistic plots
- Plots to visualize t-distributed stochastic neighbor embedding and principal component analysis
- Plots to visualize anomalies quantified via autoencoders

All these visualization and inference approaches are model agnostic; they can be applied to any of the surrogate models listed in the subsection 3.1 titled "Prediction." Many of the more computationally intensive techniques have already been applied to the example datasets for each mission and come preloaded within the ME Dashboard to provide a starting intuition about the data. For instance, by examining the precomputed variable importance plots, users may learn that the speed of the submarine was the most influential factor within the context of a specific simulated mission. They may use the partial dependency plots to identify the general trends speed has across different missions and what minimal average speed is required to meet a threshold of mission success. A submarine design can be tailored to meet that speed while then being modified with other considered influential attributes. This modification can happen using counterfactual analysis, which is a process by which users can identify the minimal change in inputs that yield a desired output via multiobjective optimization. It can be done in a constrained fashion wherein some of the inputs are not permitted to change. Within the context of the ME Dashboard, this technique allows analysts to determine if major improvements in mission success can be achieved within the local tradespace and, if so, what changes need to be made to the inputs' variables. It can help guide analysts on how to reach a desired threshold measure of effectiveness given a specified starting point.

Figure 3 shows an outlier investigation window of the ME Dashboard, where outliers are highlighted in the histogram of autoencoder reconstruction errors for a mission. The selected points are described in a summary table and automatically highlighted in a scatter plot and SHAP plot.



Figure 3. Outlier investigation window screenshot of the ME Dashboard.

3.3 OPERATIONAL CONTEXT

The ability to generate and visualize predictions can be an incredibly powerful capability, but it can also prove damaging and lead to faulty conclusions if lacking the requisite background knowledge. Without a strong understanding of the underlying assumptions of the simulation inputs that generated the data in the first place or how a measure of effectiveness is calculated, the user has the potential to misunderstand or misrepresent the outputs to decision makers.

To help remedy this issue, the Dashboard comes with briefs for each of the corresponding missions (Figure 4). Partial dependency plots are drawn for a selected attribute that is categorical in one mission and continuous in another mission. Study briefs detailing the development and results of a constructive simulation for each mission are hyperlinked in a table that also shows correlations of attributes shared between missions. These briefs are generated by the analyst as part of their customer deliverables and are typically provided as a static analysis in PDF format upon the study's completion. They detail how the higher fidelity model was constructed, what intelligence sources and environmental inputs were used, what tactics were employed, what third-party dependencies and collaborations existed, and what insights were gained from running the simulation. Complementing the briefs are graphics associated with each mission mapped on a globe to provide a less intensive, but faster review of the operational context of the mission. By incorporating these elements into the Dashboard, users can understand the underlying implications of the predictions they are generating.

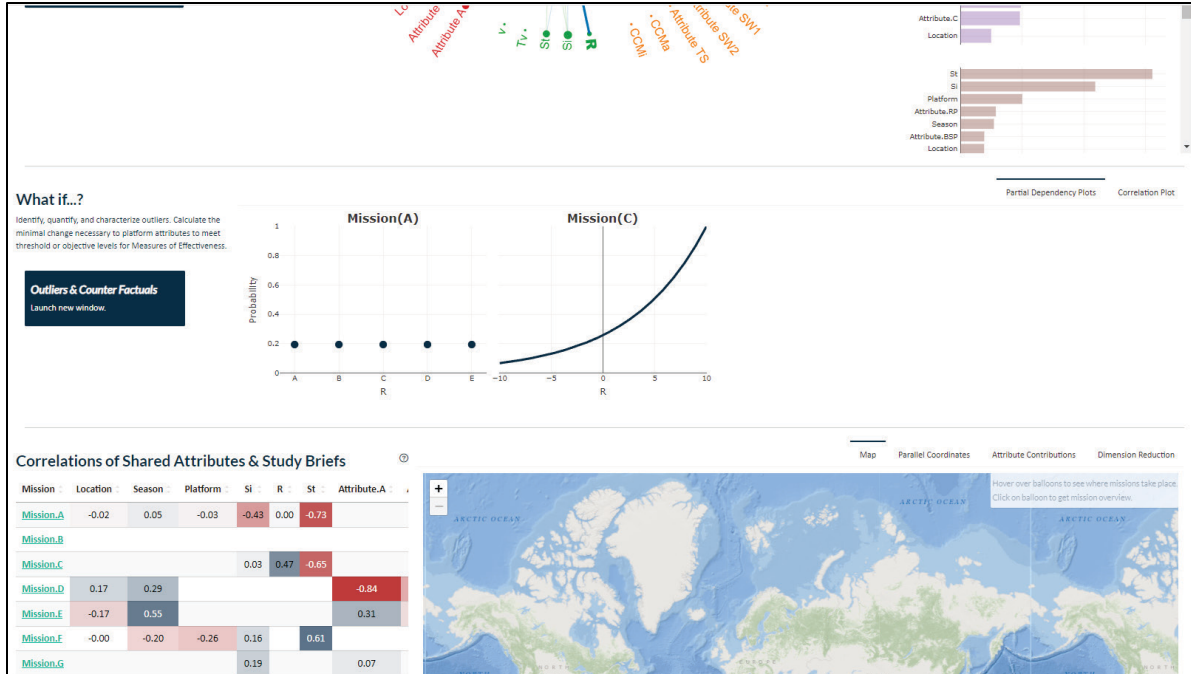


Figure 4. Screenshot of the ME Dashboard landing page.

3.4 APPLICATION AND MAPPING MODELS

Returning to the scenario in the Section 1 “Introduction,” suppose there are three missions that can be used to assess the efficacy of a submarine design—Mission A, Mission B, and Mission C. Each requires a unique set of tactics that stress different aspects of the submarine’s build (information that is detailed in the briefs housed in the ME Dashboard).

For each mission, a high-fidelity model is used to simulate combinations of relevant inputs and generate a metric of interest (i.e., some form of probability of success). This is represented in Figure 5, where the authoritative model (high-fidelity model) ingests the measures of performance (relevant inputs to the high-fidelity model) as well as contextual information to produce a measure of effectiveness. Once the simulations in the high-fidelity model are complete, the outputs are used to generate three briefs and three separate surrogate models (one for each mission) to comprise the surrogate model set housed within the ME Dashboard.

The predictions resulting from the surrogate models are used to explore the tradespace by generating various explanatory visualizations that highlight the sensitivities present in each mission and quantify the marginal impact of each attribute. The platforms with many torpedoes and higher speeds tended to do well in Mission A, whereas platforms with lower radiated noise and lower speeds did better in Mission B. While these insights help guide the architects in their conceptual submarine design, more information is required to help establish a concrete link between their designs and the impact on mission effectiveness.

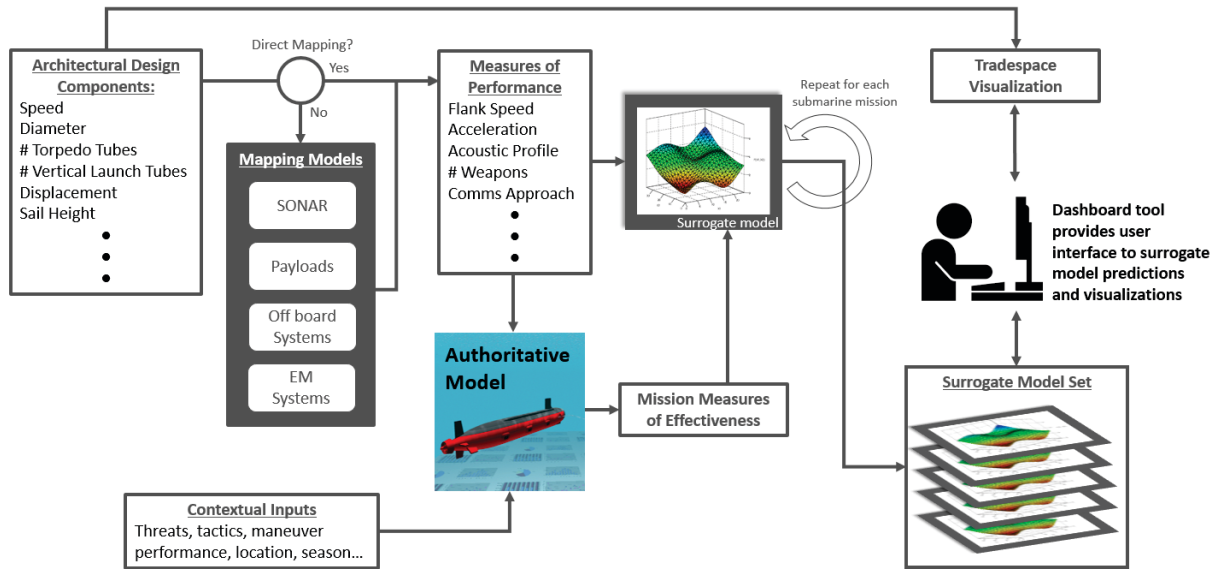


Figure 5. Submarine concept evaluation flowchart.

To establish this link, one first must recognize that many of the simulation inputs are measures of performance rather than architectural design decisions and do not map directly back to the architect’s plans. For instance, submarine speed is not an architectural component but rather a calculated byproduct of its design. Submarine propulsion, along with diameter and other drag-related features, combine to produce a speed. Similarly, the number of weapon tubes in conjunction with storage and launching systems dictates the weapon salvo rate of the submarine. These features require mapping models to translate the effect of design decisions on measures of performance that can then be ingested by the surrogate models to generate measures of effectiveness (or mission success).

Figures 5 and 6 depict the process of how a third party is required to convert architectural components into surrogate model inputs. In the previous example, a payload’s working group may act as the mapping model needed to convert the number of torpedo tubes and launching systems (which fall within the architectural design components of Figure 6) into a salvo rate that can be used by the simulation. This salvo rate is a measure of performance that can be used to generate a prediction of mission effectiveness with the surrogate model.

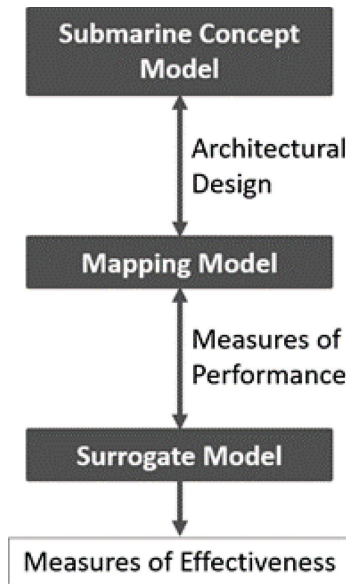


Figure 6. Information flow from architect to analyst.

These mapping models are often exceedingly complex, physics-based models that have been developed over years and maintained and improved by highly trained domain experts. The subject matter expertise of these domain experts is leveraged to convert architectural design components into inputs to these physics-based models that can then generate the measures of performance used by the surrogate models.

To further illustrate the importance of a mapping model, consider a submarine’s methods of communications with external entities. While getting more information can be beneficial in terms of directing a search for a threat platform, it also puts the submarine at risk of being detected by surface/air assets and can reduce its search rate due to slowed speeds and changing acoustic profiles in the water column. There is both a cost and benefit to increasing and decreasing the frequency of communications. In the context of a high-fidelity constructive model that simulates the submarine in a mission requiring communications between platforms, the frequency is simply represented as a scalar value dictating how often the submarine receives information; there is no direct link to its architectural components. A mapping model is required to translate this scalar value to communication subsystems, such as different types of masts. Different masts can have major design implications and force the architect to include different technologies into the conceptual submarine. It is through these mapping models that a submarine subsystem defined by architectural means can be linked to measures of performance (a communications cycle) that can then be ingested by the surrogate models to produce a measure of effectiveness.

After generating the predicted measures of performance through this translation process, analysts can assess how well a design can complete the three different missions by uploading its characterization into the ME Dashboard and generating predicted measures of effectiveness. How well a conceptual submarine does in each mission is largely contextually dependent—all non-design attributes like location, season, or threat speed were varied to yield a spectrum of

mission effectiveness for each mission. Analysts can visualize how a single design fares against competing designs that have previously gone through this process if it can be considered an outlier and improvements can be made. Furthermore, they can use the optimization tools present in the ME Dashboard to find counterfactuals. For example, if a specific design was expected to have a 63% probability of success in Mission B and the threshold architects were attempting to meet was 75%, they can identify the minimal amount of change necessary to characterize the design that will yield a 75% success rate.

Given this information, the analyst must then coordinate with the architects and subsystem experts to translate these measures of performance back to architectural features of the conceptual submarine design (see Figure 1). It is in these translations (the mapping models) where much of the difficulty lies and conflicting assumptions across organizations operating these tools are possible. Much of the success of the mission-effectiveness attribution process depends on this translation process and is still in the development process.

4. CONCLUSION

By relying on surrogate models to quickly evaluate thousands of possible conceptual submarine designs while bypassing the slow revisitation of the authoritative model, analysts can eliminate low-performing designs and, with the aid of mapping models, attribute varying levels of mission success to different subsystems. A vital step in this process, the mapping models necessitate a communicative, iterative pipeline between architect and domain experts to produce the measures of performances required by the ME Dashboard, where analysts can perform the following:

- Conduct sensitivity analysis to identify high-gradient areas in the tradespace or areas that meet minimum thresholds of performance across different missions
- Identify what attributes are common in order to be successful in multiple missions
- Identify the highest performing scenarios to see if/when specialization matters

All three aspects help the architect and analyst identify the locally optimal submarine design that yields the highest military utility given cost and building constraints. Of course, this approach is not limited to submarine design. Such benefits are transferable across industry, as these techniques are germane to applications that rely on high-fidelity, slow-running models or physical, real-world tests that are expensive and difficult to coordinate. Using these techniques, organizations can make the intelligent choice on how to focus their efforts for further analysis and investment.

Internal

Code(s): 1033 Corporate Research and Information Center (CRIC)

External

Defense Technical Information Center (DTIC)

Total: 2