2021 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
MODELING & SIMULATION & SOFTWARE (MS2) TECHNICAL SESSION
AUGUST 10-12, 2021 - NOVI, MICHIGAN

A HIGH-PERFORMANCE DATA TO DECISION SOLUTION FOR ALL ECHELON PARTICIPATION IN ARMY GROUND VEHICLE PPMX

W. Glenn Bond, PhD[1], Andrew Pokoyoway[2], David Daniszewski[2],
Cesar Lucas[2], Thomas L. Arnold[1], Haley R. Dozier, PhD[1]

[1]US Army ERDC Information Technology Laboratory, Vicksburg, MS
[2]U.S. Army Combat Capabilities Development Command Ground Vehicle Systems Center, Warren, MI

ABSTRACT

As the Army leverages Prognostic and Predictive Maintenance (PPMx) models to migrate ground vehicle platforms toward health monitoring and prescriptive maintenance, the need is imminent for a pipeline to quickly and constantly move operational and maintenance data off the platform, through analytic models, and push the insights gained back out to the edge. This process will reduce data-to-decision time and operation and sustainment costs while increasing reliability for the platform and situational awareness for analysts, subject matter experts, maintainers, and operators. The US Army Ground Vehicle Systems Center (GVSC) is collaborating with The US Army Engineer Research and Development Center (ERDC) to develop a system of systems approach to stream operational and maintenance data to appropriate computing resources, collocating the data with DoD High-Performance Computing (HPC) processing capabilities where appropriate, then channeling the generated insights to maintainers and operational decision makers where this decision support will have the greatest impact. The team has accomplished proof of concept or prototyping for several foundational components of the system and has demonstrated the effectiveness of combining data analytics with high-performance computing on large data in discovering and developing PPMx models.

Citation: W. Glenn Bond, Andrew Pokoyoway, David Daniszewski, Cesar Lucas, Thomas L. Arnold, Haley R. Dozier, "A High-Performance Data to Decision Prototyping Solution for All Echelon Participation in Army Ground Vehicle PPMx", In Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS), NDIA, Novi, MI, Aug. 10-12, 2021.

1. INTRODUCTION

The Army has been conducting ground vehicle data collection and analysis efforts to develop increased analytical capabilities that support maintenance and sustainment for many years. Larger efforts to leverage collected vehicle data occurred over the past decade through multiple Condition Based Maintenance Plus (CBM+) programs, and involved collecting operational data from thousands of vehicles and aggregating this data with additional sources of information, such as maintenance records [1]. These programs have had several focus areas, which included performing advanced analyses to provide enhanced vehicle usage, health, and reliability information, along with further supporting automating maintenance activities and reporting. Such advanced analyses yielded increased awareness and insights, highlighting opportunity areas for the Army to enhance its maintenance and sustainment approaches and increase its readiness. In particular, the development of Prognostic and Predictive Maintenance (PPMx) models demonstrates great potential to increase safety while improving Reliability, Availability, and Maintainability.

The large-scale Army ground vehicle CBM+ efforts equipped populations of vehicles with government developed Digital Source Collectors (DSCs) at Army bases spanning numerous operational environments and missions. The DSCs collect operational data broadcasted on the vehicles CAN bus (e.g., vehicle speed, engine speed, transmission gear, accelerator pedal position) in addition to diagnostic trouble code (DTC) information, as well as cumulative parameters stored in the ECU (e.g., total miles travelled, total engine hours). DSCs start collecting data when the vehicle's power is turned on, and store the recorded data until a technician downloads the data.

Once downloaded, the files are converted to the Army Bulk CBM+ Data (ABCD) file format, which utilizes the NASA's Common Data Format (.cdf) file format. These files are then uploaded to production analytics systems where they are stored and processed to generate metrics that provided enhanced vehicle information. This operational data is combined with corresponding maintenance information for each vehicle, coming from the Global Combat Support System (GCSS) Army database, and enhanced or modified by technical specialists supporting the units. Together, these data sources have supported advanced analyses that provided stakeholders at all echelons with actionable information related to vehicle status, usage, health, and maintenance assessment. This actionable information has been provided to stakeholders through emailed, automated reports as well as through web accessible dashboards.

A primary goal of the CBM+ programs and the PPMx effort has been to reduce maintenance and sustainment costs of platforms while increasing readiness. Major areas of benefit have been identified in several Cost Benefits Analyses (CBAs) conducted for certain CBM+ programs. Noted opportunities range from a general understanding of needed improvements that would provide value, such as reducing the high number of "No Evidence of Failure" (NEOF) rates, to more specific identified pursuits, such as

- Increasing the period between vehicle overhauls
- Adopting usage-based lubrication changes and scheduled servicing
- Enhancing diagnostic and predictive capabilities to enable failure reduction

Additional analytics derived benefits include

- Quantifiable cost savings
- Enhanced maintenance and sustainment support
- Injury prevention and risk reduction
- Increased asset monitoring and tracking
- Automation of maintenance and data collection processes
- Reduced mission failure rates
- Increased availability
- Logistic support

Previous programs have expended tremendous effort to collect and provide accurate data for Army ground vehicle analysis, providing the foundational requirements, processes, and tools for an end-to-end, useful data analysis framework. Implementing this framework on the scale of Army ground vehicle programs has generated a substantially large data population of sufficient quality to support development and testing of useful techniques and studies. Given the size and dimensionality of this data, past CBM+ efforts related to advanced analysis have been limited by available computational resources in terms of both processing and storage. Investigating and developing models and visualizations using large populations of high-rate, time-series data is an iterative and complex computational process.

The increased use of data analytics-based solutions in this largely serial computing effort, particularly the deployment of low-cost, open-source software for statistical and machine learning algorithms, has further increased the need for computational resources and storage capacity. Use of this data involves all phases of the traditional data science workflow, from data acquisition, cleaning, transformation, and exploration, as well as feature generation, model building and tuning, and model diagnostics and accuracy determination. DoD High Performance Computing (HPC) resources, particularly when leveraging High-Performance Data Analytics (HPDA) practices, accelerate all phases of this cycle.

Collocating the datasets with this processing power has yielded significant speedup. For example, converting all 1.07 million CDF files ion the dataset to JSON format was done in slightly less than 24 hours using one Onyx compute node. Similarly, all dataset files were rolled up into 4,482 individual histories, one for each vehicle, in one hour using 40 Onyx compute nodes. The latter, highly parallel situation, represents a 15,000 times speedup over previous efforts where files were downloaded from a webserver and individual histories generated serially. The ability to process the entire dataset quickly allows ERDC and GVSC scientists to investigate previously intractable problems, such as mining time-series data for trends that may indicate future faults, failures, or maintenance.

Migrating large datasets such as these to HPC assets will continue to demonstrate such clear benefits as workflows delve deeper into massive data. These large datasets can be leveraged by HPC resources to reveal models for real-time monitoring, visualizations for operations and maintenance, or warnings and reports for logistics awareness, and much more. Processing and storage can be scaled to fit the development stage and problem requirements. Finally, predictive models and workflows generated using data from past time intervals can be validated by comparing predicted events with occurrences from subsequent intervals for the same vehicle. HPC and cloud assets will work hand-in-glove to help move ground vehicle maintenance into the predictive and preemptive roles.

2. DATA TO DECISION

US Army ground vehicle platforms are so ubiquitous as to demonstrate an extreme of conditions cited in a 2014 Defense Acquisition University report concerning Life-Cycle Costs (LCC) [2]. The report finds that where a given fleet is large and the life-cycle long, Operations and Sustainment (O&S) costs make up the majority of LCC over the other three phases, research, investment, and disposal. Many ground vehicle platforms also enjoy comparatively low research costs, further increasing the portion of LCC represented by O&S costs. Because PPMx algorithms can significantly reduce O&S costs while simultaneously increasing RAM metrics, utilizing these models has tremendous potential to reduce total LCC.

Safety and mission success are an even greater priority for Army ground vehicle platforms. PPMx models will disrupt current paradigms by supporting go/no-go and continue/return decisions, prioritizing tasking of motor pool assets, or simply making the asset and its supply chain more reliable [3]. In a multi-domain environment, where mission success might enable objectives on multiple layers, these issues are even more critical than their elements suggest.

To maximize these returns, two conditions are essential. First, the application of models and AI/ML techniques and workflows to as many systems as possible will increase overall positive outcomes. Second, rendering the insights from these tools to the appropriate user or agency with as much time as possible to act upon the intelligence allows maximum gain in terms of better RAM and decreased costs. Stated succinctly, any useful system must find as many insights as possible and convey them to points of use as quickly as possible for immediate decision support. This section demonstrates how these investigations have identified the following principles:

- Apply new models and train them on ever-increasing amounts of data
- Collocate data and processing to maximize data throughput
- Develop and apply cutting-edge models and techniques to every system possible
- Forward insights and statistics to users and agencies for immediate decision support

The following subsections highlight how various elements of the HPDA system demonstrate proof of concept and prototype solutions, following these principles. The elements presented here demonstrate proof of concept for a production system to move data through workflows, pushing insights and visualizations to the responsible agencies, at scale.

2.1 Data on DoD HPC

DEVCOM-collected ground vehicle data comprises several Terabytes of data on DoD HPC. Several newer operational datasets are available for similar analysis as well as complementary data like weather or maintenance data. CAN bus data is collected every second by the DSC, so the dataset has become a multi-billion row volume. Files are organized in a tree structure by vehicle subtype, Vehicle Identification Number (VIN), and year of operation. Filenames within these directories contain information on subtype and VIN, then time and date, and file type. One file of type “daily” exists for every operational day for each vehicle. Challenges for this scheme include:

- Data curated separately from processing requires transfer before processing begins, adding time and cost to computationally intensive processes
- Data size for use cases following more than a small number of vehicles or more advanced models requires the computational power of HPC
- The need for fleet-wide query makes querying file-by-file inefficient

Data cleaning and transformation, as always, is a significant portion of the time and labor involved with developing any workflow. Cleaning, imputation, and aggregation are necessary prior to any visualization or model development. Documentation of these steps is also essential for validation, verification, and reproducibility. As discussed above, files downloaded from vehicles are formatted for efficient data transmission. Analysis activities benefit from transferring the individual data into a database system where more targeted, specific data subsets can be assembled. Database systems facilitate, for example, restricting analyses to specific vehicles or to specific operational locations or time-frames, but require several orders of magnitude larger storage, making them inefficient for transmission. The ABCD files from the initial dataset expanded roughly 15 times when translated to JSON format and approximately 20 times for database ingestion. The change from opening all ABCD files to the more familiar SQL based queries, however, aids in reducing analyst data preparation time while decreasing query time. This advantage is multiplied on a shared storage system as the conversion and ingestion processes need only occur once while the benefits are seen repeatedly. Further, we can accomplish common cleaning, imputation, and aggregation tasks during ingestion, providing faster access with greater commonality and established provenance.

Current Database Management Systems (DBMS) include a virtual machine-based PostgreSQL server system with TimescaleDB plugin and a container-based MongoDB server, both with accompanying databases on HPC attached storage. Both DBMS systems answer SQL-like queries, moving the results directly from files housed in the HPC-linked Center Wide File System into HPC compute node memory. The selected data can be accessed in memory by algorithms run on the compute nodes to train, test, and validate models or for several other purposes like exploratory analysis. Tests are ongoing to optimize the speed and ease with which each system addresses queries from within exploratory, model training, and visualization scripts. We have shown that production systems will need to query massive data quickly and transfer large subsets of it into systems with the memory and processor power to mine it for knowledge discovery, indicative and predictive trends, and model training and simulation. These systems have demonstrated the preliminary ability to transfer multiple gigabytes of data into memory in tens of seconds at the beginning of jobs run on DoD HPC assets.

The result of SQL-like queries is a dataframe holding only the data elements needed for the process at hand. This reduces the amount of memory used, compared to direct use of the daily files. Once dataframes are populated in system memory, the script can begin the computationally intensive work in earnest. For exploratory analytics, the DoD HPC *iLauncher* application provides a web-based interface with support for numerous development environments and tools, including *RStudio* and *Jupyter Notebooks*, and use of any desired *Python* package or *R* library to augment the capabilities found in the standard libraries. Analysts can begin a session, moving large query results into memory in a matter of moments, then manipulate the dataframe repeatedly, repeating steps until results are achieved.

Some problems may be addressed on one compute node, from an interactive interface. Multiple techniques are also available to decompose data and distribute it across several compute nodes, leveraging as many cores as necessary to mine data, run machine learning, and train sophisticated analytic models on large data. Beyond clustering packages and Message Passing Interface techniques, such tasks can be organized by the Galaxy Simulation Builder into workflows to explore hyperparameter analyses and other complex operations on large data. The tools available on DoD HPC combined with these data curation prototypes point to scalable systems developing models that will answer questions from all life-cycle phases.

2.2 Rapid Model Development

Much exploratory work has been accomplished already on GV data. Further development will be expedited greatly by the Rapid Model Development Environment. This is a virtual environment that can coexist on a user's HPC account and local machine. The environment:

- Makes commonly used, essential development libraries immediately available and any version conflicts or dependencies can be resolved and the analyst freed from managing this aspect by themselves.

- Allows Git access from HPC, enabling version control cooperation between the developer's local machine and HPC account. As model development proceeds this will aid in sharing code and modeling approaches and keeping a log of code revisions.
- Provides code snippets from existing models, including the final hyperparameter values selected and the range of model parameters investigated.
- Provides data version control, and, for models that use transformations of the raw data to create new features, increases the transparency of these modifications to data sets.

The development life cycle for the GVSC work has led to the ongoing development of several features that enable rapid workflow development and tight coupling between local computing environments and DoD HPC platforms. Python has several capabilities that allow researchers to easily create modules of code that can be shared between machines. For this project, there are numerous scripts that have been developed for all phases of the analytics workflow. These are maintained in a Git repository that, when coupled with a requirements file, helps onboard new researchers, quickly imbuing them with the same capabilities as the established team.

DoD HPC assets, as discussed above, provide the computational power necessary to mine massive data, discover knowledge in data relationships hidden there, train analytic models, and much more. The HPC environment is not, however, conducive to developing the code that runs on it. Developing and debugging workflows in the HPC environment from the beginning phases would, in fact, be a waste of HPC resources. Researchers typically perform initial development and debugging tasks on local computing resources, then migrate the code to the HPC system for final debugging and objective accomplishment. In this scenario, researchers must maintain two completely disparate code bases. The team has therefore devised a means of enabling concurrent repository access, shared between the local machine and the HPC filesystem. The system eliminates the need to maintain separate code bases while reducing the learning curve for new developers and researchers by making the developed, cumulative code base available.

There has also been much effort expended to determine the best way to maintain data integrity and provenance. Several commercial solutions exist. The team identified Data Version Control (DVC) as most promising for application to the mixed DoD HPC and local machine environment. DVC enables a git-like experience of checkpointing data as a snapshot in time, maintaining locations through *delta* updates. It encompasses the whole data lifecycle, from raw ingestion, preprocessing, feature extraction, and model development. Similar to Git, DVC provides the ability to commit changes to data and push them to a repository, with notes. The team's investigation into applying DVC across HPC and local machine environments, as was done with Git, is preliminary but promising. Like applying Git concurrently to HPC and local environments, this work will help researchers develop workflows faster and with less effort while bolstering reproducibility and provenance.

2.3 HPDA PPMx Models

As the majority of GV data is from nominal system operation, with only a small set of data from vehicles exhibiting fault or failure conditions, it is necessary to develop methods for observing changing vehicle performance. This often requires developing vehicle health metrics or an indicator of system health. This is made more challenging by the variability found in ground vehicle systems, for example the varying tolerances found in engines or transmissions, along with the high dimensionality of the data produced by these systems.

One modelling approached pursued has been a physics-informed, or physics-based data driven approach. This approach starts with deriving a performance model for the system of interest utilizing physics law or theory, such as the law of energy conservation. Direct empirical relationships are also utilized, and can be verified using the data. The derived model can then be used along with machine learning techniques in various ways. The model can be used directly by fitting the unknown parameters, such as with a regression approach, or utilizing information from the model to provide features for a neural network approach. In these cases, the data used to train the model is a subset of the population that has been classified as "healthy" vehicle data. This subsetting between "healthy" and "unhealthy" data is done as part of a separate preprocessing approach utilizing additional information sources. Training the model using "healthy" data will generate a healthy operation model, which then actual operation from a monitored vehicle can be compared to for the purpose detecting issues or predicting low health states with the vehicle. Modular HPDA workflows allow the trial of many different model and technique combinations, with similarly varying parameter sets. Testing the tuned model on remaining, reserved data allows researchers to gauge accuracy. Advantages of using physics-based data-driven approaches (beyond purely data-driven approaches) include gaining guidance for information and variables that influence the system and phenomena of interest, as well as structured features (for certain modelling approaches) that may reduce computational time when training the model. Overall, these approaches utilize knowledge and experience previously developed, increasing the chances of developing a successful approach.

Purely data driven methods, such as unsupervised learning, clustering can be used to reduce the dimensionality of the data, recognize patterns, and group similar data sets together. Using maintenance or diagnostic data, intervals can be determined from the time-series data that should contain trends discoverable by these tools. By recognizing patterns, or trends, in the data, the model can observe changes in a derived metric over time as the vehicle ages and miles accumulate (trend analysis), or as the operating environment and stresses cause faults or failures. The team has investigated a number of techniques both singly and in combination, producing significant preliminary results on limited data using

- Neural networks
- Vector auto regression
- Statistical process control
- Manifold learning

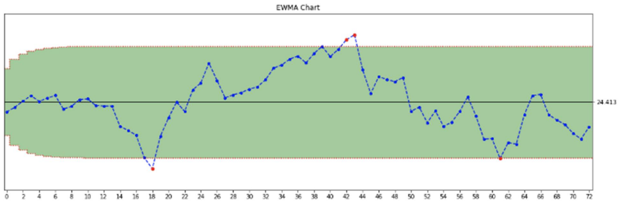

Figure 1 An Exponentially Moving Weighted Average Chart for average per operation sensor values

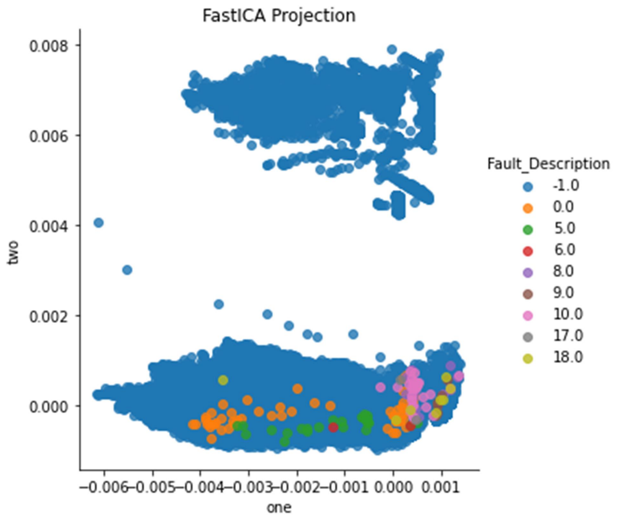

Figure 2 The 2D projection of sensor data using the Fast Independent Component Analysis algorithm. A Fault Description of greater than or equal to 0 represents sensor data that correlates time-wise to fault logs.

More methods have been explored, but these general categories have shown the promise to move forward. These will be integrated as tools in a smaller workflow for mining large operational sensor data. The data will be identified by the overall workflow as classified maintenance, fault, and failure events to intervals of operational data, providing a means of working from known events back to indicative or predictive sensor data patterns. The purely data-driven approach offers the possibility of discovering predictive and indicative sets of sensor readings, corresponding to operating regimes, without subject matter expert involvement or physics-based reasoning, expanding the number of models the system can explore and validate.

2.4 Rapid Dashboard Development

Dashboards developed from model and query driven dataframes solve use cases and answer exploratory questions for stakeholders from every part of the platform environment, such as

- Maintenance (field to depot)
- Operations
- Logistics
- Sustainment
- Manufacturing
- Design (acquisition, modification)

The ability to rapidly explore different visualizations and analytic results that are geared to a wide variety of users is another benefit of HPDA. Detailed graphs and tables that are created for an engineer are not appropriate for an end user concerned with overall fleet readiness and performance or perhaps engaged in some phase of logistic support. Having a large amount of data readily available as input to different dashboard designs allows for the rapid creation of a wide variety of user-focused output. Since stakeholders from different roles have the most complete knowledge of their own day-to-day information needs, it is more productive for them to specify what aspects of a display or output are useful for their dashboard. Providing this sort of interactivity will not only provide powerful, adaptable information sources to make platform subject matter experts more efficient, it will educate them as to insights available from the data generated by their platform. Observation of use patterns and use case requests will generate new, novel use cases for decision support HPDA.

3. FUTURE WORK

The solutions discussed in section 2 are in the prototype and proof-of-concept stages. They establish several promising capabilities in terms of potentially closing many of the gaps identified in section 1. Solutions like the ability to quickly query operational and maintenance data for an entire group of vehicles followed by rapid analysis of trends in that data have been demonstrated. Concepts for custom display of the results of these analyses, based on user role, have also been shown. Production systems based on these investigations will discover knowledge and provide unprecedented answers and insights for all stakeholders and all echelons. Besides expanding results from section 2, much more benefit can be derived from exploring how to deploy models and how to move data from the platform to computational resources and back out to organizations and units that will leverage the decision support provided.

The data-centric environment supports more, however. When developed, it will serve as a platform for further advancement of PPMx associated techniques and practices that will both stand on their own and work with other system components to move ground vehicle maintenance further toward the preemptive paradigm. The network of tools being established by this team represents the kind of information sharing and cooperation where feedback from those using the tools continually spawns an improved next iteration whose techniques and practices were not imaginable to the previous generation.

Section 3.1 discusses benefits that can be realized by moving processing out to the point of data collection. Sections 3.2 and 3.3 discuss the future of analytic models and model development and the potential to:

- Leverage collected data to discover new knowledge
- Support new life-cycle management paradigms for operations, maintenance, logistics, acquisitions and more
- Revise deployed systems through active feedback
- Move problem solving capability out to platform communities

3.1 Edge Processing Research

The size, weight, and processing power necessary to deploy PPMx models at unit level or onboard DoD platforms varies with the type of model deployed and the desired use. For example, a production system may be implemented to train multiple models, develop features, tune model hyperparameters, assess accuracy, and visualize results. Having a large computational capacity, comparable to that used for research on DoD HPC for the initial investigation, is essential in this case. Once models are trained, the computational requirements are generally significantly less, permitting deployment of models on-platform or using common, low-cost computing assets, such as a tablets or electronic control and monitoring modules installed onboard, laptops or tablets used by maintainers, or powerful but small server systems deployed with individual units. While not yet implemented, the potential to deploy trained models on-platform or for use by the operator during the mission, or the maintainer immediately after, can provide near-real-time assessment of vehicle health and allow for quickly addressing developing problems before the problem progresses.

The repercussions of such a system of systems are enormous, affecting much of the operational phase of the platform's lifecycle. At the vehicle level, indicators and predictors of fault happen in real time, allowing

- Go/no-go decisions tailored to mission urgency
- More conditions identified before the asset is tasked, increasing mission completion and safety
- More diagnostic information for the mechanic, leading to faster repair and return to mission capable status

At the unit level, maintainers and operators have the information available to

- Order parts before they are necessary and in time to schedule maintenance at an opportune time
- Access constantly available and updated information regarding which specific assets should be tasked, based on probable safety, success, and cost
- Prioritize maintenance based on maximizing safety and readiness
- Access best practices and information available from other units

Stakeholders further up the chain can receive data from edge servers at an opportune time for the communications network because immediate concerns will be addressed locally. This data can be addressed by

- Analysts and researchers verifying model accuracy and revising as necessary
- Logistics and support can evaluate supply needs and shortfalls due to transportation needs or manufacturing concerns
- Fleet planners can evaluate the lifecycle plans for the current platform and the development needs of the next platform

Although there are a number of steps to accomplish before such an overall system might provide the above benefits, research can begin to lead its development using existing data and resources. Onboard processing models, currently under development, can be tested as if deployed onboard using abundant operational data stored on HPC assets. The same data cache can be used at a larger level to simulate unit operations. On-platform and unit level computational model and software deployments can be further tested in terms of size, weight, and computational power using scalable HPC assets. Conventional computational nodes can be employed, or a multiple GPU node can simulate a deployed mini-HPC asset. Several levels of processing power can be modeled at fairly low cost before any system is specified for trial deployment.

3.2 Model Development

As Machine Learning methods continue to be refined and developed, they often appear first in open-source code libraries. The HPDA environment allows trial of the utility of these new methods on Ground Vehicle systems in a secure environment. Once the newer methods have been proven or older methods refined, these can become part of an analytical workflow applied to any of the massive datasets in storage to leverage the computational power of DoD HPC, mining massive ground vehicle data for indicative and predictive trends and patterns.

One promising new area involves the using of General-Purpose Graphics Processing Units (GPGPUs). GPGPUs are extensively used in image processing and with Deep Neural Networks (DNNs). Potential model applications include use of RNN, or Recurrent Neural Networks, and CNNs, or Convolutional Neural Networks. The former applies to time series data while the latter is most commonly used for image processing.

The training of DNNs takes significant computational resources, as the many layers of the deep network are both complex and require significant iteration to train. As there are several different DNN architectures, exploring the usefulness of the varied architectures on GV data also requires significant computational resources with access to GPU nodes on HPC essential. Deployment will require significantly less in terms of computational resources.

Another model gaining wide industrial acceptance is a digital variant of the physical system, commonly referred to as a "digital twin". The model simulates normal system operation in areas of operation not directly observed, for example when the vehicle load changes or an engine control approach has been updated. Digital twin models allow low-cost, low-risk exploration of operating regimes or environments for which little actual data exists. The simulated data obtained from the digital twin can also be used to train, update, or increase the performance of machine learning models originally built on observational data alone. Digital twin models can provide insight toward health assessment, operation in new or hazardous environments, or low cost experimentation with new configurations, and more.

3.3 Cooperative Access to Data and Models

Timely access to data, insights and visualizations for stakeholders means many different things, depending on the role of the stakeholder. The common element, however, is access to insights that inform decisions in time to make a difference in the operation, maintenance, supply chain, or lifecycle planning of assets or entire platforms. Cooperative access to data and models provides useful tools to those most familiar with day-to-day issues, giving them support for decisions in time to act for the benefit of the asset and those operating and maintaining it.

Access to platform data for academic and industrial partners means allowing access to data and tools for the purpose of collaborating in the creation of the tools, techniques, and technologies mentioned in the previous paragraph. Access must be secure to protect the sensitive nature of the data.

Computational prototyping can satisfy the needs of agency, contractor, and academic access while widening participation to include participants without HPC experience. Proof of concept exists at ERDC for secure web interface providing

- Data access and query capabilities plus basic analytics for all participants
- Existing models answering specific questions based on custom parameters and selected data
- Access to the Rapid Model Development Environment for fast, custom model and query combinations developed by non-analyst stakeholders
- A complete, web-based development environment for researchers and analysts

This combination of tools promises to reduce the cycle time needed to explore new ideas, present and assess results, and create output and visualizations for consumption by multiple users. Computational prototyping promises the sort of secure, distributed environment where stakeholders possessing operational knowledge and real use cases can collaborate with solutions providers and researchers from government, commercial, and academic concerns.

4. CONCLUSIONS

So far, GVSC and ERDC analysts have adapted workflows from existing, serial computing investigations to the parallel environment with great success. Simply adding the high-performance component to these workflows has expanded what could be done, delivering the progress of several years within months. Components and proof-of-concept for more tools were developed to reach far beyond this vision, however.

Rather than touch every file for every query, database management systems were developed to both answer the need for fleet wide query capability and to do so from processes running on HPC compute nodes. This opened the way to speedup realized before problems of the newly parallelized workflows were begun.

New tools for HPC were tested using real Army ground vehicle data and use cases. iLauncher was deployed extensively for both exploration and preliminary investigation of the data. DoD’s implementation of Singularity was leveraged to develop several useful containers, extending the tools available on DoD HPC.

New concepts are under investigation. Analytic methods are being optimized and combined in workflows to increase accuracy. New means of access to data and processing are planned and ready for investigation. New means of sharing the deeper insights and visualizations gained by HPDA workflows under development will bring actionable information to stakeholders who need it in time to act upon it. The confluence of these methods and concepts will provide a new dimension to ground vehicle analytics beyond the basic business intelligence currently offered.

4.1 Leveraging Data for the Platform

The data sources currently used are primarily CAN bus data, of SAE J1939 data. This data exists to allow on-board controllers: engine, brake, and transmission control modules, to communicate essential information for proper GV operation. Having this data stored and easily available allows for easily testing the value of additional data sources, which can then be proposed for inclusion on-platform. The ability to rapidly explore new models permits use of data fusion approaches, such as use of NOAA weather data to obtain a surrogate for ambient air temperature, which can increase model accuracy, implying this data element has high value for future vehicle designs.

4.2 Moving Processing to the Platform

Currently, all data collected on-platform is transmitted to back-end servers for processing. This fits initial model development and analysis but once models have been developed and refined it is desirable to transform significant amounts of processing to the on-platform or at-edge components, reducing the amount of bandwidth and networking requirements.

Processing functional data at the edge of deployment will yield:

- Reduced communications
- Insights on location, before data transfer
- Opportunistic, intelligent data transfer
  - What data is necessary
  - When networks are less congested
  - To the appropriate recipients
- Faster alerts and reports
  - Actionable insights in time to act
  - Tailored to the user

HPDA trained models run on motor pool based computing resources against data from local vehicles:

- Prioritize which vehicles to task
- Forecast local O&S needs
- Optimize RAM metrics
- Predict local fleet's ability to perform mission tasking

Onboard processing:

- Real-time Go/no go and return/continue
- Situation-dependent Condition Indicator
  - Protect lives in battle scenarios
  - Improve safety and protect property in training scenarios

4.3 Continuous Revision

As additional data becomes available, particularly data for systems that have developed a failure, the models can be refined and improved. Hyperparameter sets can be retuned and the model accuracy rechecked. Models will be revalidated using large computational resources and secure data storage of the central, enterprise level HPDA system. Cooperation enhances the process through involving multiple organizations with different skill sets. Engineers, computer scientists, and mathematicians, for example have easy access to the data and models, which encourages exploration and enhancement of the models.

To date, maintenance data has been used sparingly, aided by DEVCOM DAC's Sample Data Collection (SDC) program, which uses field service personnel to provide more detailed information on vehicle health status and repairs conducted. Efforts are underway to link Enterprise Resource Planning (ERP) systems, such as GCSS-Army, which contains maintenance actions and vehicle status, with the on-platform data. The inclusion of maintenance data holds promise to improve model accuracy but also complicates aspects of model exploration as there are no guarantees that parts replaced were always the cause of observed performance symptoms, a phenomenon commonly known as No Evidence of Failure (NEOF). Additional data exploration and outlier analysis is often helpful to identify these cases.

REFERENCES

[1] S. Das, "An efficient way to enable prognostics in an onboard system," in *IEEE Aerospace Conference Proceedings*, 2015, vol. 2015-June, doi: 10.1109/AERO.2015.7118976.

[2] G. Jones, E. White, E. T. Ryan, and J. D. Ritschel, "Investigation Into the Ratio of Operating and Support Costs to Life-Cycle Costs for DoD Weapon Systems," *Def. ARJ*, vol. 21, no. 1, pp. 442–464, 2014.

[3] W. Glenn Bond *et al.*, "A Hybrid Learning Approach to Prognostics and Health Management for Military Ground Vehicles Using Time-Series and Maintenance Event Data," in *Annual Conference of the PHM Society*, 2020, vol. 12, no. 1, pp. 10–10, doi: 10.36001/PHMCONF.2020.V12I1.1146.