

Developing Requirements to Validate Autonomous Ground Vehicle Simulations

J. Cox, S. DeRosier, A. Howell, J. Jones, and B. Thompson

Department of Systems Engineering
United States Military Academy
West Point, NY 10996

Corresponding author's Email: shae.derosier@westpoint.edu

This paper is a reprint and was the SISE Best Paper Award winner at the 2022 GDRKMCC Conference Proceedings.

Author Note: The Capstone Team would like to thank Mr. Jed Richards, MAJ William E. Ponder, and the Engineer Research and Development Center Information Technology Laboratory (ITL), Institute for Systems Engineering Research (ISER), Geotechnical and Structures Laboratory (GSL) and the Mobility Systems Branch (MSB) for giving the team the opportunity to be a part of the Simulation Tools for Combat Vehicle Robotics (ST for CoVeR) program research and development. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract: This project developed a methodology for assessing virtual testing of the United States' military autonomous vehicle platforms. The authors conducted background research of autonomous vehicles and simulation to determine the best virtual testing methods that provided the strongest evidence of simulation performance. The team created a requirements list for the simulation software of autonomous vehicles to help drive the virtual test development and conducted statistical analysis, pairwise comparisons, and visual analysis tests to assess simulation data compared to data from physical test runs. From this study, the authors identified multiple methods of virtual testing that can assess the autonomous vehicle simulation for further development.

Keywords: Autonomous Vehicle, Simulation, Virtual Testing

1. Introduction

Simulation tools are becoming increasingly popular for evaluating models and prototypes of innovative technology. Simulation technology is being used by military developers to design and test military autonomous vehicles which use Light Detection and Ranging (LiDAR) sensors to autonomously navigate obstacles in an unconventional environment. Simulations are useful in research and development because they can provide valuable performance data of emerging technology at a fractional cost due to less resources being required for testing. Currently, the simulations for autonomous vehicles are undergoing an in-depth assessment through virtual testing by comparing the Virtual Engineering Evaluation Test (V-EET) results from simulations to their Physical Engineering Evaluation Test (P-EET) result counterparts conducted at military testing facilities. The final goal for the military developers is for the simulation to be reliable enough to predict physical performance of the autonomous systems in any environment around the world.

The authors identified virtual testing requirements and developed methodologies to assess the fidelity of the simulation for autonomous vehicles. This paper provides the methodologies for assessing virtual testing of autonomous ground vehicle (AGV) platforms and analyzes data of both virtual and physical testing runs with the goal of aiding military developers in validating and verifying AGV simulations.

2. Literature Review

The study began with a literature review focused on autonomous vehicle simulation including research of military unmanned ground vehicle platforms and the goals for its future capabilities, especially in an operational or combat environment. The research looked at uncovering and understanding current and future needs of the U.S. Army in terms of acquisition and usage of autonomous vehicles. Research in this topic field falls under the development of Next-Generation Combat Vehicles which is currently the United States Army Futures Command's second highest priority for modernization that is necessary for future readiness and Multi-Domain Operations (U.S. Army Futures Command, 2022). Literature reviewed explored how autonomous vehicles can be integrated and utilized by the military to improve combat effectiveness. Military uses of AGVs

found in literature included weapons platforms, weight carriers, reconnaissance assets, and engineering operations (National Research Council, 2002). Lastly, the research highlighted current difficulties with the development of such complex systems. A common challenge found throughout the reviewed literature was AGVs' performance in inclement weather such as rain, fog, and other forms of precipitation as well as environmental factors like brush, tall grass, and dense vegetation (Shacklett, 2019). However, resources were the greatest challenge and barrier for autonomous system development and testing, often proving to be extremely costly and difficult to acquire (Shacklett, 2019; National Research Council, 2002). Simulations are one tool that can be implemented to decrease or eliminate resource requirements, such as those necessary for physical testing, dramatically reducing the costs for research and development.

The literature review also explored existing simulation methods of autonomous vehicles and various methods to validate and verify simulation accuracy. After creating a virtual model, it must be verified, validated, and improved to close the performance gap between simulation and reality. The literature review clearly indicated that a properly validated simulation must have a defined degree of accuracy or fidelity when compared to the physical environment it replicates (Sargent, 2013). However, no one method of measuring accuracy or an acceptable level of fidelity was agreed upon to determine the value of a simulation (MITRE, 2018; Sargent, 2013). There were common conclusions in the literature about the challenges researchers face in developing simulations. One such challenge was that producing a higher fidelity model requires greater processing speeds, which requires increasingly powerful hardware able to handle the processing load (Zhao et al., 2021). Another challenge is apparent when virtually simulating LiDAR technology and its effectiveness in detecting obstacles in obscured environments. Many physical tests of LiDAR sensors show that systems report false alarms when attempting to operate in rain, snow, or fog (Jokela, Kutila, & Pyykonen, 2019). A major challenge in the virtual testing of AGVs is replicating the environmental effects and simulating the LiDAR technology response in order to accurately predict vehicle performance. Lastly, the literature review focused on military simulation technologies, specifically Virtual Autonomous Navigation Environment (VANE) and Autonomous Navigation Virtual Environment Laboratory (ANVEL). These two software packages are the primary testbeds for the military AGV simulation. VANE is used to create a virtual environment which replicates the physical world using survey-grade LiDAR (Jones, Priddy, Horner, & Peters, 2008). ANVEL is a physics-based system that is used to recreate the autonomous platform and allow it to navigate in VANE (Holland, 2014).

3. Virtual Testing Requirements and Validation Methods

3.1 Simulation Requirements

The simulation validation study began with the development of a list of the functional and design requirements for the simulation. The requirements worked as a guide to better determine which test metrics provided the greatest indications of the system's performance. Figure 1 shows an example requirement as well as the method used for developing the requirements. The requirements were drawn from military autonomy developer's current and historical work on the simulation and discussions with the research team. First, the requirements were identified and defined in a clear, concise statement. Next, the requirements were classified by type. The researchers documented the originating source of the requirement either from literature or military requirements. The team then developed the thresholds, defined as the minimum performance criteria, and the objectives, defined as the goal or ideal performance criteria. Both thresholds and objectives were developed by referencing expert knowledge of military autonomy developers, derived from literature, and gleaned from available test data which is explored further in the next section. Finally, researchers identified which simulation metrics related to the requirement that would provide an indication of how well that requirement was being met.

| ID | Definition | Type | Threshold | Objective | Measurement/Metric |
|----|---|-----------------------|---------------------------------------|----------------------|--|
| 2 | The simulation for the autonomous vehicle shall replicate the same environmental effects and damage from weather and battlefield conditions on the LiDAR sensors as the physical environment. | Functional, Allocated | Rocks/Shrapnel Friction Force +/- 10% | Small Arms Munitions | Number of detections wanted, unwanted, and missed. LiDAR performance compared to previous tests |

Figure 1. Example Requirement for Simulation Assessment

3.2 Assessment Methods

To understand how to validate the simulation, it was necessary to categorize and breakdown each type of virtual test, show how they relate to the different components of a simulation, and how the tests are used to provide validation evidence. Figure 2 depicts a validation matrix of methods developed by the authors and military autonomy developers to test and validate the simulation performance. The validation matrix allows developers to evaluate the performance of a single test or compare multiple test runs (shown in columns in the matrix) across all elements of virtual testing (shown in rows in the matrix). The elements of virtual testing include the virtual environment, sensors, vehicle autonomy algorithms, and vehicle behavior. The Environment is the virtual environment created by the simulation system. Sensors refers to the LiDAR, GPS, and camera systems the virtual platform uses to navigate within the virtual environment. Autonomy is the artificial intelligence system that is gathering and interpreting data from the sensors and using it to navigate the environment. Behavior is defined as the actions of the autonomous vehicle as it navigates the environment. Across the top row are broad qualitative and quantitative analysis areas or “bins” for analyzing and validating the virtual test. The table shows how specific tests can validate all the identified simulation components and analysis areas. In the following sections, the authors demonstrate several of the proposed methods from the validation matrix including qualitative measures of visual comparisons of route performance and sensor cost map plots, and quantitative statistical measures to evaluate and compare physical and virtual AGV test run data.

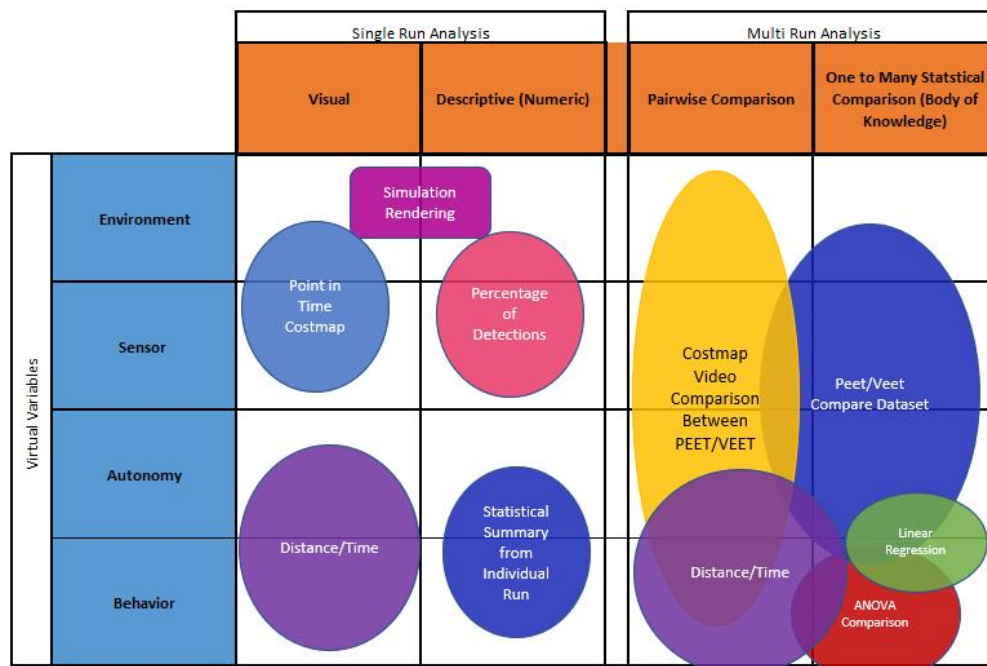


Figure 2. Validation Matrix of Methods for Simulation Assessment

4. Simulation Assessment Tests

4.1 Statistical Analysis

Statistical analyses were conducted to evaluate the environment, sensor, autonomy, and behavior across single and multiple runs by using a combination of two different software navigation algorithms, two routes, and four measurements (total distance traveled, average speed, percent time moving, and total moving time), to determine if a statistical difference existed between the measured performance of the physical and virtual testing. Ideally, no significant difference should be observed between the testing environments, thus validating the autonomous performance in both environments. Therefore, the null hypothesis was that the physical and virtual simulations were similar. The data violated the Shapiro Wilks normality test assumptions due to the small sample sizes for both physical and virtual test run data. As a result, the authors conducted non-parametric statistical tests, specifically Mood's Median Test. A significance level of $\alpha = 0.05$ was used for all statistical tests.

The Mood's Median statistical test results in Figure 3 show no significant difference for each combination of algorithm and route excluding algorithm B's average speed on route 1. A significant difference in the median average speeds for algorithm B suggests one of two things. First, the autonomy for algorithm B may have had an issue with controlling speed or decision-making abilities. Second, the virtual simulation needs to be improved to represent algorithm B's driving speed more accurately. It is hard to draw conclusions from route 2 because there was no data for algorithm B's performance on this course. Overall, the statistical analyses suggest that initial virtual testing is performing similar to the physical testing. Military autonomy developers can use the current test statistics taken from the small sample of runs to analyze if the values improve as more data is collected.

| Algorithm | Route | Total Distance Traveled | | Average Speed | | Percent Time Moving | | Total Moving Time | |
|-----------|-------|-------------------------|---------|---------------|--------------|---------------------|---------|-------------------|---------|
| | | Chi-Square | P-Value | Chi-Square | P-Value | Chi-Square | P-Value | Chi-Square | P-Value |
| A | 1 | -0.14825 | 0.882 | -0.70419 | 0.481 | -1.8161 | 0.069 | 1.5196 | 0.129 |
| B | 1 | -0.073561 | 0.941 | 2.354 | 0.019 | -0.73561 | 0.462 | -1.6149 | 0.106 |
| A | 2 | -0.20412 | 0.838 | 0.40825 | 0.683 | -0.40825 | 0.683 | 1.4289 | 0.153 |

| | |
|------|---|
| | Fail to reject the null hypothesis. Physical and virtual are similar. |
| <.05 | Reject the null hypothesis. Physical and virtual are significantly different. |

Figure 3. Virtual and physical test comparison by algorithm and route, Mood's Median statistical analyses

4.2 Route Overlays Analysis

The team also developed a method to visually compare autonomous vehicle route navigation performance. Figure 4 shows a latitude and longitude representation of the routes taken by each algorithm on each course. Orange lines denote the physical path taken, while blue depicts the virtual path. Figure 4a shows algorithm A's performance on route 1. The route overlay shows that the autonomous vehicle performed well, completing the course while only going slightly off course on the initial stretch. In comparison, algorithm B followed route 1 more precisely, with lines of absence in the data. This may indicate that the simulation is more geared towards accurately representing algorithm A compared to algorithm B or that perhaps algorithm B's data output ability failed to record its GPS data. Figure 4c shows algorithm A's route performance on route 2. Clearly, both physical and virtual routes are inconsistent and incomplete, possibly indicating a weakness in the vehicle's autonomous capabilities when the course gets harder. Even in Figure 4d, the autonomous vehicle did not perform the same on two physical test runs. This can be viewed with a second physical test run in green overlayed on the original algorithm A, route 2 physical run. With the low amount of data collection for route 2 and its apparent randomness, it is better to use route 1 in assessing the virtual simulation capabilities until additional test run data is available.

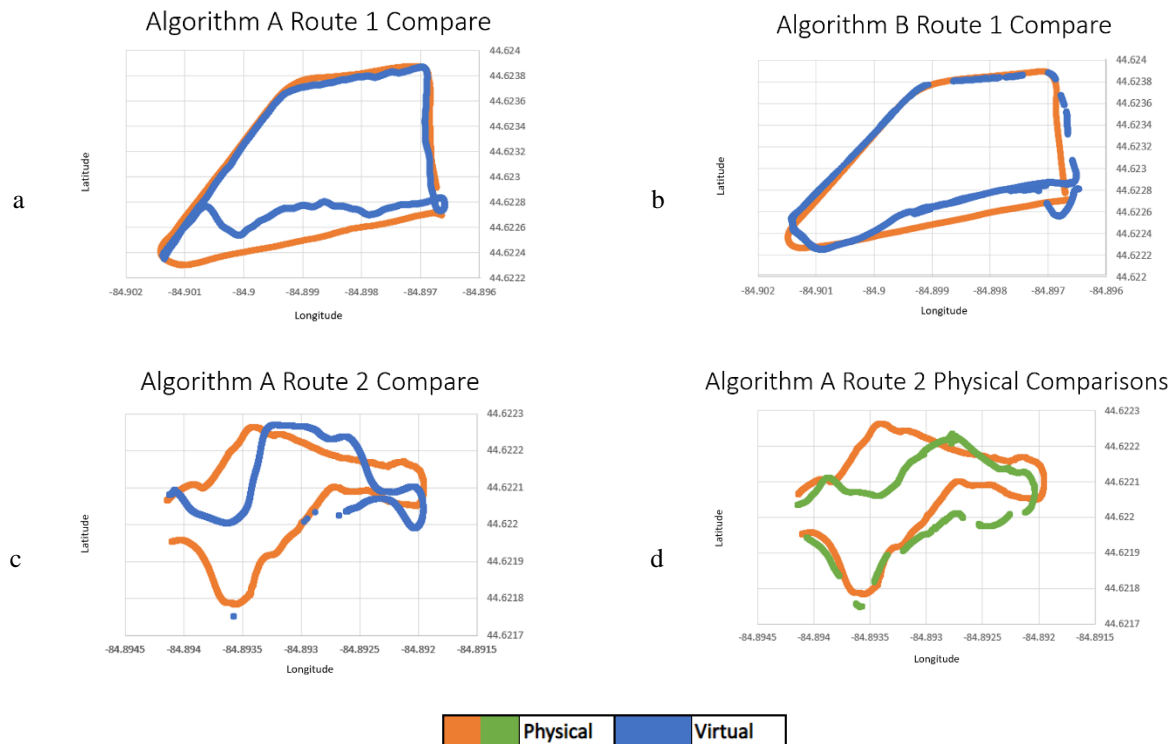


Figure 4. Physical (orange and green lines) vs Virtual (blue lines) Route Overlays a) Algorithm A Route 1 Compare b) Algorithm B Route 1 Compare c) Algorithm A Route 2 Compare d) Algorithm A Route 2 Physical Comparisons

4.3 Cost Map Analysis

Cost maps represent another virtual test tool to assess the simulation's performance, specifically the environment and sensor performance. Cost maps are a snapshot of identified obstacles detected by the autonomy based on LiDAR sensor output while it navigates an environment. Military autonomy developers can use cost maps to help show that AGVs are detecting the obstacles accurately. Figure 5 contains cost maps of the LiDAR sensor from physical and virtual test runs. These cost maps were created by using RStudio to stitch together the collected data. Lighter areas of the cost maps indicate where the system has determined it to be low cost, or low risk to travel while darker shades indicate areas of high cost, or high risk of travel has been identified. The vehicle autonomy then chooses the least costly path to travel. These snapshots are collected roughly every half second during a test and can be scripted together to observe the run from the LiDAR's perspective. In all images, the vehicle is in the center and sends out a signal to its surroundings.

The first two images are from two different physical test runs on the same route. While the cost maps are not identical, they do both correctly identify an obstacle, which in this case is a building located at the physical test site, represented as an outline of the box in the lower middle portion of the cost map. The virtual cost map depicted on the far right is from a virtual recreation of the same course; however, it looks drastically different from the physical cost maps. The reason that there is such a difference between the virtual and physical cost maps is unknown at this point. One potential reason for this difference is that the virtual vehicle may be seeing these obstacles but is having a difficult time projecting it in a similar view. Developers can use this method to identify discrepancies between the physical and virtual test data and evaluate changes made to the simulation environment and sensors.

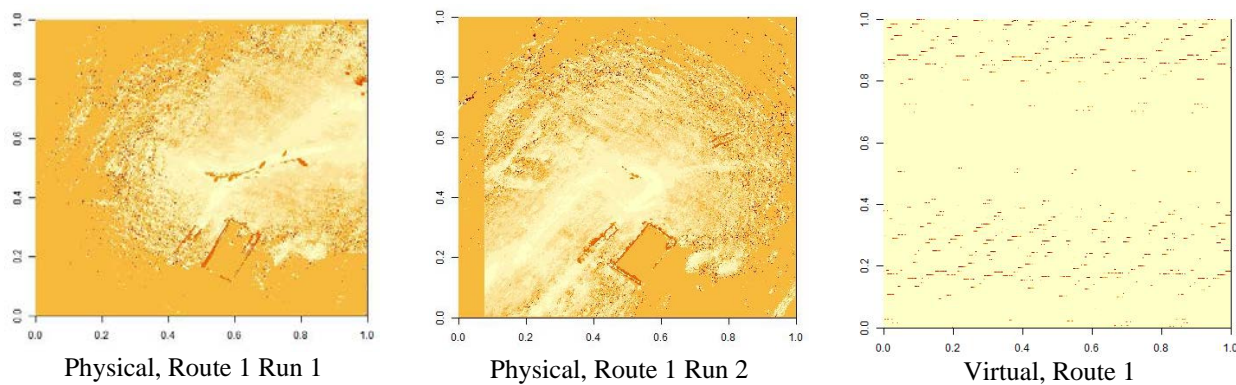


Figure 5. Algorithm A, Route 1 Cost Map Comparison

4. Conclusion

As shown through each of the validation tests, the existing autonomy simulations would not be a valid substitute to physical testing because the simulation technology currently provides incomplete data and the number of samples is small, therefore, cannot be completely assessed. To become a valid test substitute, the authors found that examining the four metrics above, total distance traveled, average speed, percent moving time, and total moving time, provide a holistic summary analysis of each run and can be used to determine the success and quality of each virtual test run. This is supported by the statistical test results shown in Section 4.1. Cost maps and route overlays are also valuable tools in visually examining the behavior of the autonomy algorithms, both for single runs and multi-run comparisons. Using visual comparison is a quick, effective method for determining the success of a run during testing. As development of the simulation for autonomous vehicles continues, the authors recommend conducting similar validation tests with an increased number of samples of both physical and virtual test runs.

5. References

- Holland, M. (2014). Sensor Deployment Analysis with the Virtual Autonomous Navigation Environment. Engineer Research and Development Center. <https://www.erd.c.usace.army.mil/Media/News-Stories/Article/476622/sensor-deployment-analysis-with-the-virtual-autonomous-navigation-environment/>.
- Jokela, Maria, Kuttila, Matti, and Pyykonen, Pasi. (2019). Testing and Validation of Automotive Point-Cloud Sensors in Adverse Weather Conditions. *Applied Sciences*, 9(11), 2341. <https://doi.org/10.3390/app9112341>.
- Jones, R.A., Priddy, J.D. Horner, D.A., and Peters, J.F. (2008). Virtual Autonomous Navigation Environment (VANE). *Earth and Space* 2008. [https://doi.org/10.1061/40988\(323\)114](https://doi.org/10.1061/40988(323)114).
- MITRE Staff. (2018). Verification and validation of simulation models. The MITRE Corporation. Retrieved September 27, 2021, <https://www.mitre.org/publications/systems-engineering-guide/se-lifecycle-building-blocks/other-se-lifecycle-building-blocks-articles/verification-and-validation-of-simulation-models>.
- National Research Council. (2002). Technology Development for Army Unmanned Ground Vehicles. Washington, DC: The National Academies Press. <https://doi.org/10.17226/10592>.
- Sargent, R. (2013) Verification and Validation of Simulation Models. *J Simulation*, 7, 12–24. <https://doi.org/10.1057/jos.2012.20> and <https://link.springer.com/article/10.1057/jos.2012.20>.
- Shacklett, M. (2019). LiDAR: Promise and Challenge in Autonomous Vehicles: What happens in Detroit doesn't stay in Detroit when it comes to LiDAR. *Point of Beginning*, 45(2), 8+. <https://link.gale.com/apps/doc/A607581245/PPMI?u=west10360&sid=bookmark-PPMI&xid=25ed7ef0>.
- United States Army Futures Command. (2022). Who We Are: Cross Functional Teams. *Army Futures Command*. <https://armyfuturescommand.com/cft/>.
- Zhao, Jian, Li, Yaxin, Zhu, Bing, Deng, Weiwen, and Sun, Bohua (2021). Method and Applications of Lidar Modeling for Virtual Testing of Intelligent Vehicles, *IEEE Transactions on Intelligent Transportation Systems*, 22(5), 2990–3000. <https://doi.org/10.1109/tits.2020.2978438>.