Study Design to Assess Autonomous Mobility of the Experimental Unmanned Vehicle (XUV)

Barry A. Bodt and Ann E. M. Brodeen
U.S. Army Research Laboratory

Unmanned ground vehicles (UGVs) will provide scout functions for our forces on the future battlefield. A current study of the Experimental Unmanned Vehicle (XUV) seeks to demonstrate autonomous mobility. The study design attempts to balance military operational and development-related technical concerns, multiple sites, restrictions on randomization, and resource constraints to maximize the information quality and content resulting from testing. The principal design follows a split-split plot scheme, and this answers most questions. However, some additional testing is necessary to address subordinate issues. As in most field trials, trade-offs must be made between the statistical ideal and the practical reality. Considerations run the gambit of pooling, confounding, and nesting, and also the questions of fixed versus random factors and the advisability of using a portion of data in two separate analyses. The panel of experts to which this paper is presented will be asked to respond to these trade-offs with any guidance they have regarding the design or the analysis.

More information from Barry Bodt …

Enclosed in this e-mail is a briefing on this design that the I delivered on 23 October at Fort Indian Town Gap, PA before representatives from NIST, ARL, and General Dynamics. With the caveats mentioned in the briefing, the design as it stands has been fairly well received.

One specific area in which the panel might provide guidance is pooling. Douglas Montgomery's 2002 fifth edition text on Design and Analysis of Experiments suggests on page 536 that what terms given up to pooling might be determined by first testing the significance of the term—perhaps with a high alpha level, say 0.25. I don't have a feel for an appropriate approach, but some pooling will be required because we simply do not have enough degrees of freedom in the denominator for many of the tests. Some rough power computations, convince me I probably should have double the number of replicates I can afford in this test.

Another issue that has been brought up is the fact that the XUV/team factor is inseparable in the present design. That was a concession we made. The problem is that if we include XUV and Team as separate factors, to support randomization we may have to shuttle teams back and forth to the two test courses. Logistics cost us time and consequently runs. So this was a trade-off we agreed to. Still, if there was a way to block or in some other way cleverly arrange the design so that Team and XUV could be separated, it would be nice. Any suggestions along that line would be most welcome.

Ultimately, a second site will be tested. The way this is set up now, I would be driven to a split-split plot design. However, I need to be careful on the analysis. For example, the test course—even with specified difficulty level—is really nested within site. I hope not, but it is even possible that technical operators will be, at least, somewhat different the second time around. Alerting me to any landmines in the analysis would be very useful.

Another issue is the borrowing of runs from Tech-T1 in the principal experiment (12 to be exact) for use in another comparison involving soldiers and night conditions. That's not ideal either and if weather conditions change for the Tech-T1 runs from the principal experiment to Excursion 2, I am only going to have to settle for an unbalanced design. Randomization restrictions are not well accounted for either. Any thoughts on this are also welcome.

Currently, manned HMMWV runs occur on a separate day than the principal experiment involving the XUVs. Given the manned HMMWV's are the baseline, this is not ideal either. There is some thought about lengthening the work day to accommodate some manned runs in the morning of say, days 8 and 9, and perhaps in the evening on days 2-7. What we play against is fatigue on the part of the safety crew and the test administrators. That could also influence the end test result.

The responses in this study are geared toward autonomous mobility. The robot is supposed to carry out a mission, traveling to certain GPS designated points in three mission distance configurations. The real issue is, generally, how much help did it need to get there? The way it is addressed is in terms of the number of operator interventions necessary, the percentage of time the robot is truly autonomous, the number of emergency stops (e-stop) that safety invokes to protect the equipment, operator workload to keep track of the robot, and the percent of mission distance completed. It is possible the robot will get stuck or e-stopped and will not be able to continue.

There are other questions that you could focus on and I certainly don't expect you to address all of these in the time we will have. Moreover, you may choose one of the other questions I allude to in my briefing or another that has completely escaped me.

My final comment is to let you know this is a real test and the program is a very big deal in the Army robotics community. Never before have we tried to collect data that addresses military-operational considerations and never before has a test of this size been attempted. And, although the schedule is tight, testing does not begin until the 2nd of December, although a shake-out test at Fort Indian Town Gap will occur in mid November. NIST has been contracted to administer the test, with us involved, and so provides test directors, terrain assessment, and some personnel.