Flight Test Methods for Unmanned Aircraft

Chapter 20: Small UAV Testing
Chapter 21: Highly Autonomous Systems

Major David Zyga, United States Marine Corps
47123 Buse Rd.
Bldg 2272 Rm 255
Patuxent River, MD 20670
UNITED STATES

David.d.zyga.mil@us.navy.mil

Chapter 20: Small UAV Testing

ABSTRACT

Small UAS (SUAS) present a different set of testing challenges to the aviation test and evaluation professional. Due to a wide commercial market, modern militaries are often turning to commercial-off-the-shelf (COTS) SUAS solutions in lieu of developing bespoke mission-oriented systems. SUAS tend to be inexpensive and expendable, but there remain safety risks that require different mitigations than manned aircraft. Most SUAS do not require specialized aviation training, so the stakeholder community is not limited to aviation professionals. Test program management must balance the requirements for defendable and repeatable results, with the low-cost and limited life cycle of the assets.

20.1 INTRODUCTION

Small UAVs present the tester with unique challenges. In many ways, a small UAV has the same relationship to large UAVs that insects do to birds…when you get below a certain size, the rules change. The famed aerodynamic impossibility of a bumblebee flying comes to mind. Test methods and approaches must be altered. However, the fundamentals of safe test execution and mission-relation of test results remain the cornerstones of sound test practices. If anything, they become more important with small UAVs, for many of the quantitative test methods become much harder to execute in the smaller scales.

20.1.1 SUAVS ARE DIFFERENT

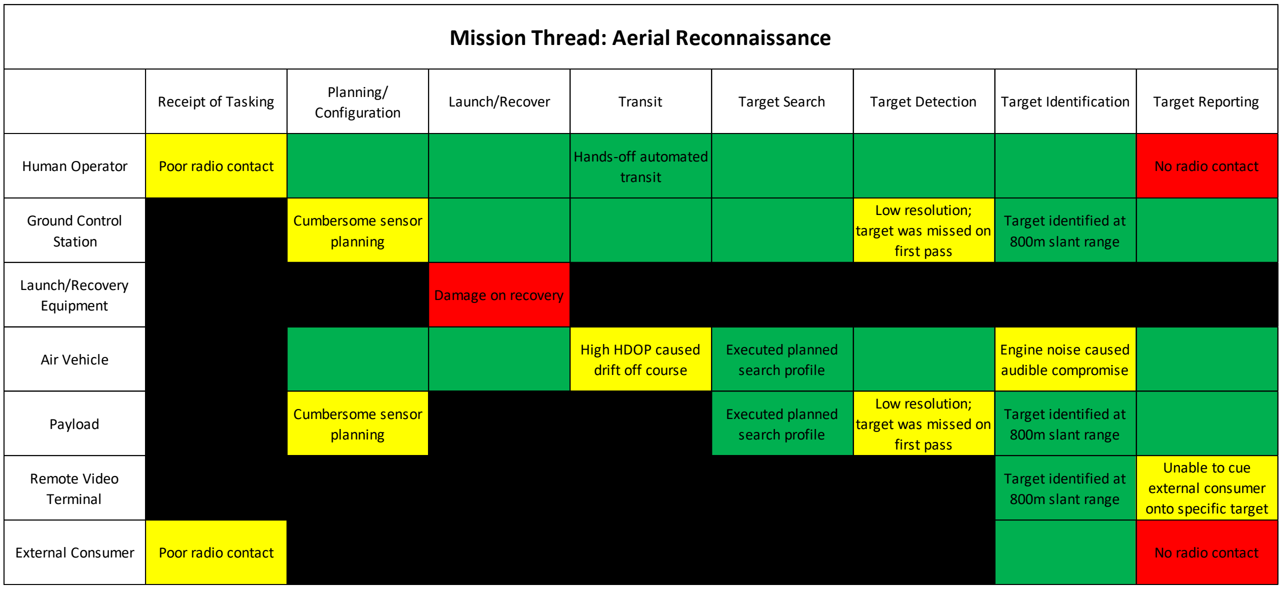
Unlike their large UAV brethren, small UAVs have sufficient commercial and consumer use cases to fuel a commercial UAV industry. This industry is driven not by government requirements and design specifications, but by rapid advancements in a variety of technologies: material manufacturing, energy storage, propulsion efficiency, novel sensors, machine intelligence, information processing, and data exchange. While the defense sector was once a driver for development in these areas, it is now in a position of trying to maintain pace with private industry and academia.

20.1.2 SYSTEM OF SYSTEMS

It is common for legacy testers to define the system under test entirely by the boundaries of the aircraft, and this approach may also have some merit with the largest unmanned aircraft. However, as the airborne component diminishes in size, more functionality is necessarily offboarded to ground support equipment.

- Propulsion migrates from the powerplant to launch and recovery equipment (LRE)
- Navigation and data processing migrate from the mission computer to the ground control station (GCS)
- Data dissemination flows not from the aircraft, but from the GCS
- Multisensor platforms become multiplatform sensors

Any test regime for a UAV should approach the system under test as a system-of-systems, and consider the vernacular UAS. The additional (non-flying) components of the system serve not only as enablers, but often critical nodes in the mission thread that require deliberate evaluation within the mission context.



Mission Thread: Aerial Reconnaissance

	Receipt of Tasking	Planning/ Configuration	Launch/Recover	Transit	Target Search	Target Detection	Target Identification	Target Reporting
Human Operator	🟨 Poor radio contact	🟩	🟩	🟩 Hands-off automated transit	🟩	🟩	🟩	🟥 No radio contact
Ground Control Station	⬛	🟨 Cumbersome sensor planning	🟩	🟩	🟩	🟨 Low resolution; target was missed on first pass	🟩 Target identified at 800m slant range	🟩
Launch/Recovery Equipment	⬛	⬛	🟥 Damage on recovery	⬛	⬛	⬛	⬛	⬛
Air Vehicle	⬛	🟩	🟩	🟨 High HDOP caused drift off course	🟩 Executed planned search profile	🟩	🟨 Engine noise caused audible compromise	🟩
Payload	⬛	🟨 Cumbersome sensor planning	⬛	⬛	🟩 Executed planned search profile	🟨 Low resolution; target was missed on first pass	🟩 Target identified at 800m slant range	🟩
Remote Video Terminal	⬛	⬛	⬛	⬛	⬛	⬛	🟩 Target identified at 800m slant range	🟨 Unable to cue external consumer onto specific target
External Consumer	🟨 Poor radio contact	⬛	⬛	⬛	⬛	⬛	🟩	🟥 No radio contact

Figure 20-1: Sample System-of-Systems Mission Evaluation Matrix.

20.2 SPECIAL SAFETY CONSIDERATIONS

Small UAVs present significant safety-of-test issues. The size of the smaller UAVs, and their relatively low cost, can tempt testers into neglecting safety precautions that would normally be taken for larger manned or unmanned aircraft. It can definitely tempt program managers into pressuring testers to ignore safety issues in the name of schedule and cost. However, this can lead to serious problems. A 20-kg UAV may not be large or expensive, but hitting a house with one that went out of control will cause no end of trouble. Any mishap involving the outside community will not redound to the credit of the test community in general nor the test team in particular. Safety must not be neglected.

When possible, testers should adhere to the safety precautions described in Chapter 2 of this test manual. In particular, the requirement for dual links to the air vehicle, and a policy that failure of a single link constitutes an immediate knock-it-off criteria, should be observed unless the design of the UAV precludes this precaution.

Small UAV testers should also consider carefully the implementation of flight termination systems, or a reasonable equivalent capability. Small UAVs offer a variety of equivalent safety approaches not applicable to larger unmanned aircraft. Limiting the amount of fuel carried can keep a UAV "gone wild" on the test range. Safety tethers can be used to restrain a small VTOL UAV, just as they have many larger VTOL aircraft. And for very small UAVs, initial tests can be conducted inside hangars. An airship hangar, in particular, offers a large volume of space for initial testing, all with a protective cover that keeps weather out and UAVs in.

However, testers should also bear in mind that many small UAVs are cheap enough to be considered attrition assets. A level of risk that would be considered unacceptable for a manned aircraft program may fall well within bounds acceptable to the managers of a small UAV program. Testers working small UAV programs should discuss the level of acceptable risk to the aircraft with the program management and test squadron leadership and agree on a tolerable risk level. This discussion must bear in mind that while small UAVs may be expendable, homes in the outside community are not. This may impose boundaries on risk significantly tighter than the UAV alone warrants, or force the test team to perform testing in more remote locations.

20.3 SMALL UAV TEST PROGRAM MANAGEMENT

Managing a test program for small UAVs presents several challenges to the test team.

First, program sponsors sometimes will become unrealistic, and attempt to rush a small UAV into operational testing or even operational employment without adequate testing to ensure safe and proper operation and characterization of performance. The test team will have to persuade program managers that investing in a thorough test program will solve problems that would otherwise lead to embarrassing mishaps in the field. Once a system gets a poor reputation with operational users or with the general public, its future is often dim. Proper testing can avoid such problems.

Second, small UAVs are frequently intended for use by operators who do not have aviation backgrounds. This can present test management with several significant challenges. Best practice is to conduct testing with personnel who do have aviation backgrounds, but this may meet with resistance from program sponsors not fully aware of the safety considerations involved in UAV flight testing. As with the schedule/scope of test issue, the test team will have to clearly communicate the benefits of using properly trained test personnel.

Non-aviation-qualified operators present a third management issue, that of relating test results to the capabilities of a non-aviator. The situation is similar to testing a primary trainer aircraft – the test pilots must evaluate the aircraft not from the perspective of a trained aviator, but a student. Testers may find it worthwhile to conduct first flight, envelope expansion, air vehicle performance, and sensor testing using aviation-qualified operators, then pass the debugged system to more Fleet-representative operators for a final assessment.

Fourth, small UAS tend to have shorter flight times. Battery-powered UAVs may only have 20 minutes of flight time per battery, which does not give much time to run through a test card. The test team will have to find ways to maximize efficiency, and avoid wasted time while airborne. Range coordination, working area entry, target setup, and ensuring safety measures and observers are in place should all be conducted prior to launch. The data collection plan for each sortie should account for realistic flight times, and multiple test sorties may need to be planned to execute a single test card.

A fifth issue to be mindful of is the cost ratio of a test program to the UAV. Or, more appropriately, the cost of test should not exceed a certain percentage of the cost of *not* testing. As a general rule, the cost to test a candidate system should not exceed 25% of the program’s procurement value, or it should not exceed 25% of the cost of mission failure. For example: a $10M procurement of 100 UAVs might suggest an upper bound of $2.5M on testing activities; however, if the cost of mission failure is somewhat higher (such as a missile-countermeasure UAV), then the cost of failure may serve as a better benchmark of program value—in this case, testing should be planned for around 25% of the repair or replacement cost of the ship or installation being protected by the UAV. When scoping a test program, test managers need to understand stakeholder appetite for statistical defendability; for the preponderance of small UAVs, the stakeholders will have a high risk tolerance that justifies a truncated test program.

20.3.1 COTS TEST PROGRAM MANAGEMENT

There is a growing appetite in the defense community to maintain parity with commercial UAV technology by fielding commercial-off-the-shelf (COTS) systems. In all but the most niche applications, there is a commercial application for everything a military operator would do with a UAV. These commercial systems tend to be inexpensive, purpose-built for an analogous commercial use-case, and provide a simple method for iterative replenishment and technology upgrades every few years. It is tempting for program managers and other stakeholders to skip testing and field directly to the users; however, this approach often fails to reveal mission-related inadequacies and interoperability issues that limit its practical utility. When a system is discovered to be ineffective in the field, the users are often multiple years away from their next iterative upgrade.

While COTS products are not custom-designed to military specifications, they typically comply with industry standards that are similar enough to military standards to be sufficient for military use. The industry vendors will have conducted a level of test that satisfies consumers and civil aviation agencies throughout system development, and prior to the product offering—this significantly reduces the scope required for independent government testing. In most cases, government test teams should consider the following items as being largely satisfied by industry, and not requiring further testing.

- Flight safety
- Aircraft handling
- Datalink reliability
- Propulsion reliability
- Navigation/control reliability

In those cases that the user community has a specific requirement, such as operational availability, that should be the focus of a government test program.

While basic system functionality may not change significantly between commercial and military users, the military operations provide a different environment and mission context than the system may have been designed for. For COTS systems, it is highly advisable to conduct a limited-quantity operational test in a realistic mission environment prior to procuring mass quantities. This may even be a “fly-off” between competing systems to educate the user community on the benefits and shortfalls inherent to each system. The Information Technology community has been using this approach quite successfully for a number of years, and it has produced an iterative model worthy of replication.

Figure 20-2: Agile Approach to SUAS Program.

Key enablers:

- Stable, measureable, repeatable user requirements
- Technical evaluation of vendors' internal development and test program
- "Canned" requirements-based test design and report, crafted for ease of re-use
- Pre-planned "go/no-go" criteria for full procurement
- "Canned" user feedback/evaluation metrics and reporting mechanism during fielding period

Provided that the user requirements do not change significantly between iterations, the small UAV test program can be designed as a "canned" program to be re-used each cycle. This yields significant savings in the time to design, plan, and report each test, potentially saving months at a time. This proves extremely valuable when subsequent iterations are an upgraded model of the system from the same vendor—the test team need only focus on the "deltas" between the models.

There is a growing trend in the commercial UAV industry to standardize components of the system. Flight computers, datalinks, ground control stations, and human-machine interfaces are often re-used on different model systems, and even with multiple vendors. These similarities give military users several benefits: common training curricula, reduced time-to-train on new systems, and increased maintainability. The learning curves apply equally to the test community: a user interface that was found to be favorable on a previous system is likely to remain favorable on a current or emerging system, which potentially reduces the requirement for further testing.

20.4 FLYING QUALITIES

Small UAVs present several significant flying qualities challenges.

First, some small UAVs use classical joystick-type controls, with the operator directly commanding control surface positions. Such designs can be tested using slight variations on methods for manned aircraft. In particular, performing mission tasks and using the Cooper-Harper scale will work. It should be noted that more and more small UAVs are going toward the “fly-by-output” approach used by 4th generation large unmanned aircraft. The development of small attitude sensors for consumer products has made this advance cost-effective.

Second, the operating environment at low altitudes makes testing that requires stabilized test conditions impossible. This affects performance testing far more than flying qualities, but does impact flying qualities testing significantly.

However, testing focused on mission tasks can still be performed. Mission relation is the final test of a system’s suitability for Fleet use, engineering data is merely supplemental. And it should be noted that control systems that allow for closed-loop control of the air vehicle can be tested with existing tools such as the Cooper-Harper scale.

The third flying qualities challenge is the potential for very unconventional designs. Large unmanned aircraft, such as the MQ-4 Triton or MQ-8 Firescout, have designs very similar to manned aircraft of equivalent size and performance. With the small UAVs, this changes. The quadcopter design commonly used for recreational UAVs has no manned aviation equivalent. Some ornithopter UAVs have been flown. And approaches such as tail-sitting aircraft, which were tested in manned form and rejected as impractical, become entirely feasible as UAVs.

Quantitative testing of such unusual designs may prove difficult. However, qualitative testing should be feasible using existing test methods and a strong focus on mission relation. It’s advisable to pay particular attention to configuration and mode transition changes. Historically, both manned and unmanned aircraft using substantial changes in configuration have encountered problems in this area.

The size of some small UAVs severely limits some tolerances that are otherwise acceptable on larger aircraft. Minor vibrations, voltage fluctuations, and propeller RPM variances may not be noticeable in a Group 3 or larger aircraft, but can make a Group 1 or 2 aircraft completely unflyable. It may not be possible to physically instrument such a small aircraft to detect these nuances, so testers may find that metadata and log files provide useful information to detect and measure these adverse effects.

Figure 20-3: Open-Source Log Analyser.

20.5 PERFORMANCE

20.5.1 POWERPLANTS FOR SMALL UAVS

The test methods presented in FTM-108 are designed for jet and turboprop powered aircraft. Small UAVs frequently use reciprocating engines, or even electrical power. Some emerging systems add propane, hydrogen cells, and solar power to the list of energy sources. Test methods for aircraft using reciprocating engines are presented in Chapter 11 of the U.S. Air Force Test Pilot School Performance Phase Textbook, USAF-TPS-CUR-86-01. Testers of such aircraft may find it advisable to review general aviation performance test methods, due to the extensive use of reciprocating engines in such aircraft.

As test methods for reciprocating engines use power, not thrust, as the major test variable, test methods for such powerplants should be transferrable to electrical propulsion systems. Testers dealing with ornithopters or other exotica are advised to conduct a search of recent publications. Some papers have been written dealing with performance testing, but it is an emerging technology. Testers who successfully develop test methods are strongly encouraged to publish professional papers of their own to document lessons learned.

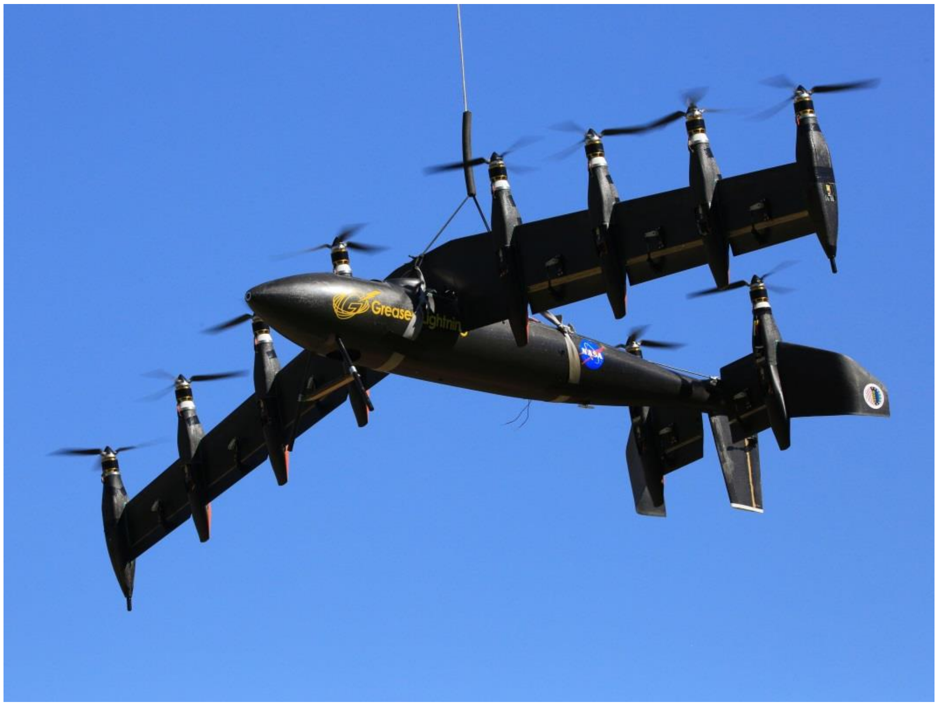


Figure 20-4: NASA Electrically Powered UAV Conducting Hover Tests.

20.5.2 ATMOSPHERIC EFFECTS

Performance testing is strongly affected by the turbulent environment in which small UAVs operate. A small UAV normally operates at altitudes of only a few hundred feet at most, well within the boundary layer of the surface/air interface. Furthermore, the low weight makes small UAVs considerably more susceptible to atmospheric turbulence. As a result, small UAVs must contend with turbulence that has significantly more effect on the aircraft than their larger, higher-flying counterparts. All this makes effective performance testing more difficult.

Some mitigation can be found by conducting testing in the very early morning, when the air will be the most calm. Further mitigation can be found by testing over water, where the lack of terrain can provide smoother air for testing.

However, testers may find themselves compelled to use more “brute force” test approaches. This is particularly true for range and endurance testing. The frequent throttle adjustments required to maintain airspeed and altitude in turbulent air makes stabilized performance testing nearly impossible. The Brute Force approach, flying for an extended period at a specified altitude and airspeed, can produce time-averaged data which is highly mission-relatable.

Testers should also exercise care with the design of the test program, which should be tailored to the operational environment and needs of the end user. Early in the planning process, they should bound the flight envelope to those operational use cases described by the concept of operations. Aircraft expected to operate close to the ground need not be tested for range and endurance at higher altitudes. Likewise, an aircraft expected to operate across a narrow speed band need not be tested at a large number of airspeeds. And quite often, coarse performance data will suffice. Given the schedule pressure on many small UAV programs, the program management will be willing to trade precision of performance data for speed in fielding.

20.6 LAUNCH AND RECOVERY

Small UAVs frequently use unusual launch and recovery methods. Hand launching, a variety of catapults, and even rocket boosters have frequently been to get a small UAV into the air. And a range of nets and cables have been used to recover small UAVs. Testers should be wary of rejecting an approach out of hand – the idea of deliberately flying a UAV so that the wing would hit a vertical cable may seem like an excellent way to lose aircraft, but the RQ-21 uses it quite successfully.

Figure 20-5: RQ-21A Blackjack Catapult Launch.

Figure 20-6: RQ-21A Blackjack Cable Recovery.

Testing such aircraft demands flexibility in both planning and execution. Close attention should be paid to stability and control during transitions from one configuration to another, and from one guidance mode to another. Figures of merit in testing will include the absence of uncommanded movements during mode transitions, and both vertical and horizontal error margins during recovery. If the small UAV is recovered by flying into a net or catching a cable with a wingtip, both the reliability of that method and the damage done should be noted. This will require careful record-keeping by the test team, as each flight will present only one data point.

20.7 HUMAN FACTORS

Human factors for small UAV control systems also warrant close attention by the test team. Even more than large UAVs, small UAVs suffer from control stations designed by people with no aviation background and poor understanding of the operational environment. When possible, the test team should work closely with the designers to eliminate deficiencies in the design phase. Where not, testers must thoroughly test and document deficiencies quickly for correction prior to deployment. Given the fast pace of many small UAV programs, early identification of deficiencies should be given a high priority.

20.8 SURVIVABILITY TESTING

Traditional aviation survivability testing generally focuses on human life preservation during an aircraft emergency, or as a result of adversary action. Life support equipment and weapon countermeasures typically dominate survivability testing efforts, because these are the aircraft's principal mechanisms to protect its human occupants from harm. Survivability equipment may even be conceptually extended to include system health monitoring to warn of a pending failure, or threat detection systems prior to adversary engagement.

It may be tempting to overlook survivability in an unmanned aircraft; however, the modern battlefield may be just as lethal to a remote pilot as it is to a human in the cockpit. The UAS command and control (C2) datalink, particularly in a small commercial UAS, is often conducted via omnidirectional line-of-sight radio frequency (RF) emissions from both the air vehicle and the ground control station. These emissions can be exploited by sophisticated adversarial forces to detect, fix, and track both the UAV and the human operator. In extreme cases, the datalink can be exploited to harvest sensor data and telemetry that reveals compromising information about the friendly force's location, composition, disposition, and intent. The datalink is the weakest link.

Both the telecommunications industry and the defense communications community recognize the need for secure data transmission and protection of the transmission source. These material and non-material solutions may be applied to small UAS as design features, field modifications, or procedural workarounds.

- Encrypted communications enhance the security of the information, but do not defend against detection.
- Operating in a saturated frequency band increases the RF clutter and reduces the likelihood of detection.
- Low power transmissions decrease the ability to be detected from long range.
- Frequency agility and spread spectrum techniques mitigate some detection risk by rapidly cycling through multiple radio frequencies and reducing power density.
- Directional antennas inhibit adversaries' ability to detect from all azimuths, and may potentially reduce the required transmission power.
- Antenna offset or retransmission may physically separate the human operator from the source of the outgoing emissions.
- Beyond line of sight (BLOS) datalinks can displace the operator to a sanctuary location outside the adversary's range.
- Short-duration or burst transmissions can inhibit the adversary's ability to detect and react to an emission in a timely manner.
- Automated flight may reduce or eliminate datalink usage altogether.

Test teams should strive to evaluate warfighting systems in the context of their intended operating concept and threat environment. While test programs may not necessarily resolve a vulnerability, there is tremendous value in simply identifying and characterizing the risk to force and risk to mission so that mitigating measures can be implemented. Small UAV program practitioners would be wise to consider these vulnerabilities when generating requirements for the next cycle of small UAV procurement.

20.9 CONCLUSION

Small UAV testing demands both technical competence and mental flexibility from the tester. They should not be underestimated. On the other hand, small UAVs allow the test team the chance to work with some very unusual design approaches, and build their professional reputations for being able to quickly and effectively test exotic designs. Above all, they offer the chance for testers to support the Fleet quickly – and there are few things more satisfying than getting a vitally needed capability deployed in months instead of years.

Chapter 21: Highly Autonomous Systems

ABSTRACT

Autonomous system test and evaluation is an emerging discipline that is being constantly redefined. There is no single authoritative definition of levels of autonomy that cleanly applies to all platforms or applications; as such, there is no single autonomous system test methodology suitable for all platforms or applications. The tester's challenge is to build a clear understanding of the expected autonomous behaviour across each task or condition as a basis for evaluating performance. Autonomous systems open up the possibility to deploy multiple UAS as a "swarm", which requires a system-of-systems approach to test and evaluation.

21.1 INTRODUCTION

Autonomy has been a feature of unmanned aircraft since the mid-1990s. 4th generation UAVs have taken both the manipulation of the flight controls and the execution of immediate-action emergency procedures out of the hands of the pilot and put them into the hands of the mission computers. However, these designs all rely on a human operator to make all critical decisions, either by commanding flight parameters such as altitude and course, or by pre-programming a mission plan or emergency procedures. Execution is autonomous, decision-making is not.

But this is changing. Experiments are underway with a new generation of capabilities, featuring a much higher level of autonomous decision-making than their predecessors.

Testing the limited-autonomy unmanned aircraft of the last twenty years presented challenges to the flight test community. Methods to persuade an autonomous guidance system not intended to fly flight test maneuvers to do so had to be developed, ways to interpret test results created, and techniques to manage multi-shift, multi-discipline test missions invented.

Highly autonomous systems present even larger challenges. The scope of testing must expand to cover the decision-making logic, as well as the flying qualities and performance characteristics evaluated in the past. And experience teaches us that a flight tester, particularly of unmanned systems, must plan well ahead. Trying to come up with test methods on the fly is a recipe for embarrassing failure.

21.2 DEFINITION OF HIGHLY AUTONOMOUS

Several definitions of levels of autonomy have been developed. Testers should be wary of letting arguments over definitions supplant test planning and execution. Experience has shown a risk of teams spending too much time and effort struggling to define terms and neglecting to plan tests.

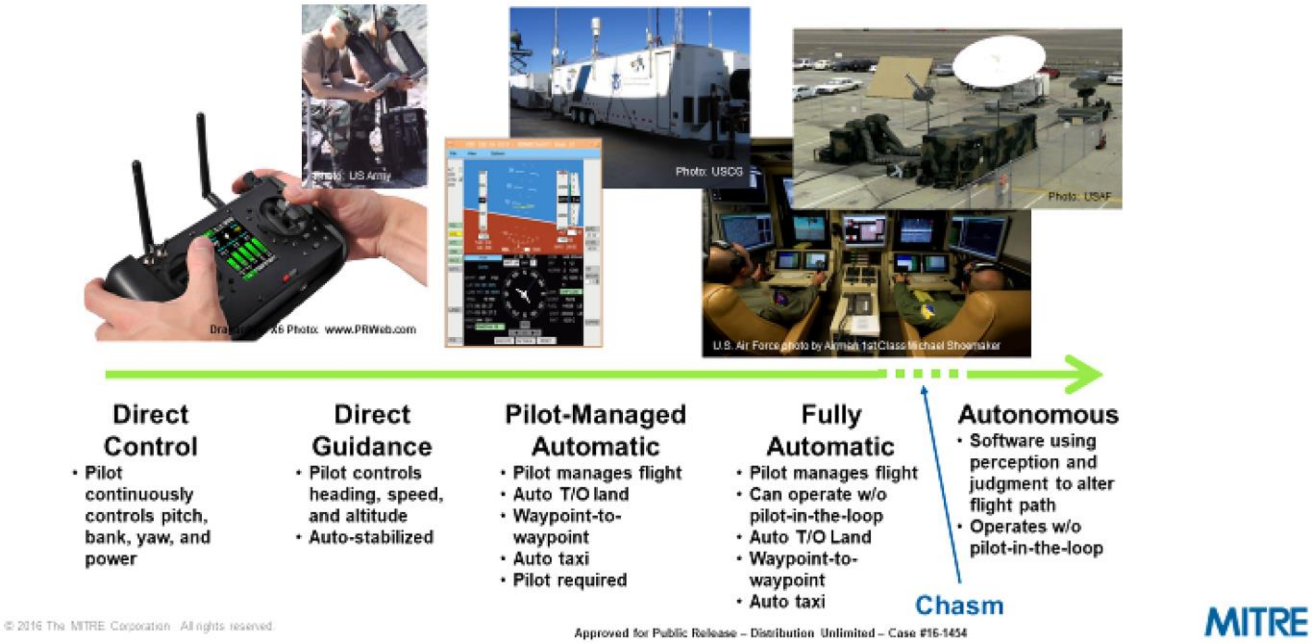


Figure 21-1: UAS Autonomy Levels (Courtesy of The MITRE Corporation).

Table 21-1: Taxonomy of Autonomy (Courtesy of Sheridan & Verplank, 1978).

Taxonomy of Autonomy		
Autonomy Level	**Description**	**Colloquial Analogy**
1	Human gives commands, with no machine assistance	Teleoperation
2	Human gives commands, with several machine-aided options	Low-level decision support
3	Human gives commands, with few machine-aided options	Medium-level decision support
4	Human gives commands, with single machine-aided option	High-level decision support
5	Machine makes decisions, with human approval	Human-in-the-loop
6	Machine makes decisions, with human veto power	Human-on-the-loop
7	Machine makes decisions, and informs human	Human-near-the-loop
8	Machine makes decisions, and informs human upon request	Human-aware-of-the-loop
9	Machine makes decisions, and informs human if it decides to	Human-unaware-of-the-loop

10	Machine makes decisions, and does not inform human	Human-not-required

21.3 PHILOSOPHY OF TESTING HIGHLY AUTONOMOUS AIRCRAFT

In developing a test approach, it pays to study history and similar systems used for other applications.

From a historical perspective, it becomes clear that highly autonomous decision-making capabilities must be considered separately from basic airworthiness and performance testing. Test methods developed for the last generation of unmanned aircraft are still applicable to testing the basic airframe. Highly autonomous decision-making should be considered an addition to the test program, not the only thing on it.

From the similar systems perspective, there are several autonomous systems in common use that provide examples of both potential and problems. While autonomous cars and their crashes garner headlines, the most familiar autonomous system can be found on the dashboard of most cars – a navigation system. Type in your destination, give it a moment to compute, and driving directions are provided, complete with an annoying voice to simulate an obnoxious back-seat driver. (This is an example of a machine giving commands, and the human relegated to the role of a mere steering wheel actuator.)

Or so the theory goes. In reality, the results have been less than ideal. Navigation systems that give impossible and even unsafe directions have become a meme. And automobile autopilots have repeatedly killed people. These systems may be autonomous, but they are reliant on operator override for safety.

Before a tester can rigorously evaluate an autonomous system, she must first have a level of understanding of what the system does and how it works. It may be helpful for testers to approach the system as a systems engineer would approach a functional allocation model: determine which tasks are conducted by which components, and assign metrics to evaluate those tasks. In some cases, success criteria for a machine is doing a task better than a human: speed, precision, and accuracy are three common quantifiable reasons to allocate control or decision-making to a machine. In other cases, the quantifiable metric is a specification that was derived from a requirement or standard. Determining the effectiveness of the system goes beyond assessing the machine in isolation; it is inexorably teamed with the human operator(s), and the combination of man and machine serve to symbiotically complete mission tasks.

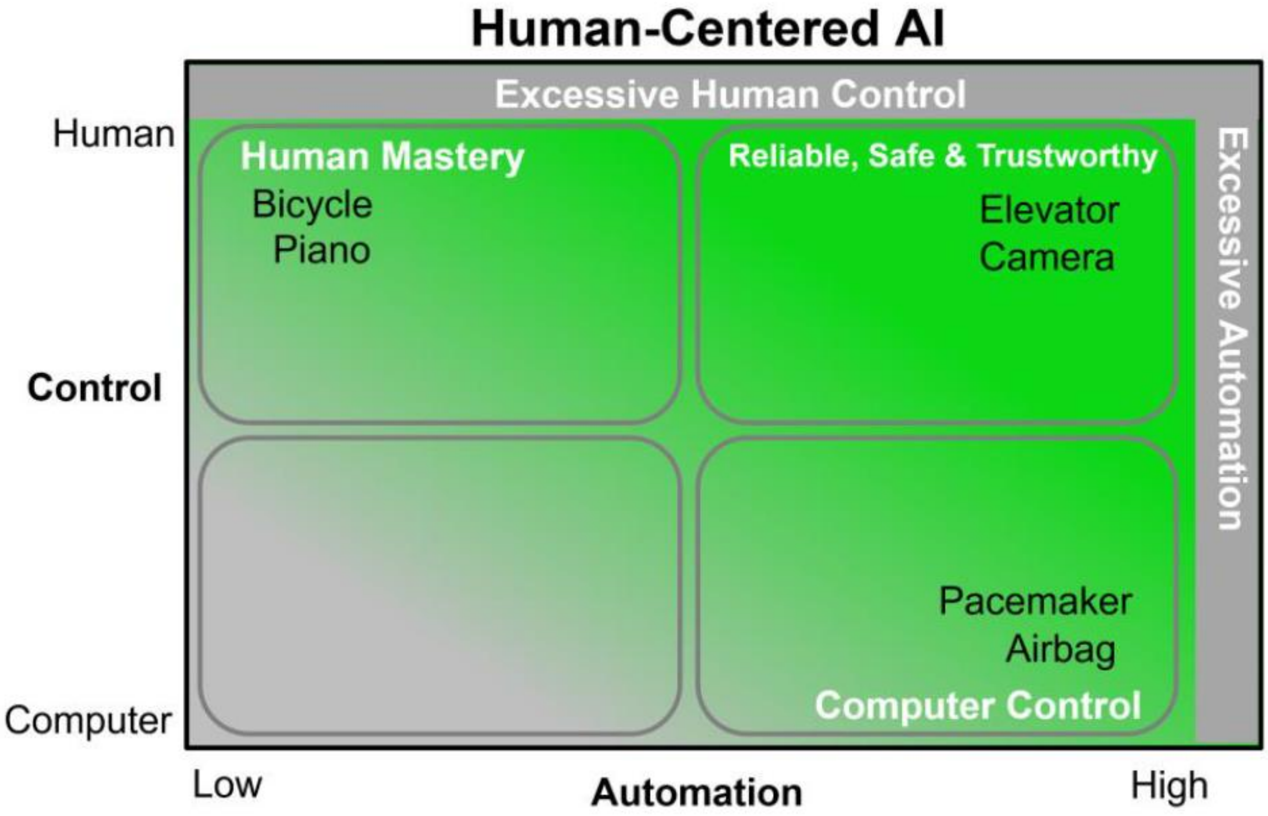


Figure 21-2: Human-Centered Artificial Intelligence (Courtesy of Shneiderman, 2020).

Table 21-2: Sample Human-Machine Teaming Evaluation Matrix.

Autonomous VTOL Landing Allocation and Evaluation Criteria			
	Human	**Machine**	**Criterion**
Landing site selection	90%	10%	P_S of depicting potential landing zone = 0.8
Landing site survey	50%	50%	Surveys site size and slope +/- 10% error
Obstacle detection	50%	50%	P_D of obstacle > $1m^3$ = 0.9
Commit to landing spot	100%	0%	< 1m horizontal error in spot selection
Approach to landing spot	0%	100%	Maintains 300 +/- 25 fpm descent from 500 AGL to 10 AGL

Touchdown	0%	100%	Maintains 100 +/- 10 fpm descent from 10 AGL to surface Touchdown point +/- 1m from intended landing spot
Ground sensing	0%	100%	P_S of weight-on-wheels detection = 0.95
Abort landing	75%	25%	P_S of autonomous abort in fouled zone = 0.8 < 1 sec reaction time to operator-commanded abort

21.4 FUNDAMENTAL TEST APPROACH

Considering the problem, it becomes clear that testing of a highly autonomous aircraft must separate the testing of the basic airframe from testing of the autonomous decision-making capabilities. A new airframe carries an immense amount of risk in itself. High levels of autonomy also carry immense risk. Combining the two into a single system without making sure they work properly alone practically guarantees a major mishap.

Another matter that is clear is that most of the testing of the autonomous decision-making capabilities will have to take place in a Systems Integration Laboratory (SIL), on the ground. A properly equipped SIL, able to fully replicate the operating environment, external inputs, and (if appropriate) other cooperative platforms, will allow the autonomous decision-making capabilities to be safely tested on the ground. Flight testing will need to be performed as a final check, but the primary testing will need to be conducted in the SIL.

Finally, flight testing must encompass more than a simple functional check. Flight testing of an autonomous system for military use must evaluate its utility for operational missions.

21.5 INITIAL AIRFRAME TESTING

Flight testing 4th generation unmanned aircraft presented many challenges. While the basic principles remained the same, there were many differences in approach required to adapt test methods for manned aircraft to unmanned platforms, especially those with autonomous execution of flight commands. In particular, it has proven necessary to incorporate a library of Engineering Test Commands (ETCs) into the guidance software to excite aircraft dynamic modes, assist in performance testing, and support safety of flight assessments.

This requirement for specialized test software will be present, and may well be increased, for the highly autonomous systems of the future. Envelope expansion and performance testing of an aircraft intended to ultimately be controlled by a highly autonomous decision-making system may require the development and incorporation of an interim guidance system restricted to autonomous execution of pilot decisions.

It's worth noting that experience has shown that any test-specific software requires approximately 18 months lead time – which, in turn, requires that the envelope expansion and performance flight test planning be started far enough ahead to determine exactly what ETCs will be required and get those needs to the software team in a timely manner.

Testers should also bear in mind that not only will ETC's be required for the air vehicle, but corresponding capabilities to command and disengage them must be incorporated into the ground station. Which means that the requirements must be clearly communicated to the team writing the ground station software. Do not assume the air vehicle and ground station software are being written by the same people, or even the same company. And attention must be paid to the way in which the aircraft returns to normal operations when an ETC is cancelled – especially an abrupt cancellation for safety reasons.

The amount of extra work involved in creating an interim guidance system will depend greatly on the approach being used for the software. A tight integration of autonomous decision-making and autonomous execution will probably demand a lot of work to substitute an interim manual system. On the other hand, a system architecture that separates the decision-making and execution functions can have a pilot-based decision-making system incorporated far more easily. It is probably prudent to make this separation in the original software design concepts.

And the test software must also support an override capability for later testing of the highly autonomous capabilities. Experience to date with experimental highly autonomous systems has shown that a "snatchback" capability, a way for a pilot to take control away from the autonomous decision-making systems, is essential to keep testing safe. Range safety personnel (not to mention the general public) will be far more comfortable knowing the "HAL 9000" flying the UAV can be overridden.

Finally, test managers should note that it is likely that initial airframe testing and tests on the autonomous decision-making capabilities will be conducted in parallel. This will affect overall test team manning.

21.6 GROUND TESTING HIGHLY AUTONOMOUS CAPABILITIES

As discussed before, the primary test environment for the highly autonomous decision-making capabilities will have to be the Systems Integration Laboratory (SIL). A classic SIL is an "iron bird", with a full set of air vehicle hardware coupled to simulators and stimulators that allow the system to be tested as if it was in flight. Partial SILs, with some systems simulated, are also frequently used. Although a SIL (or more likely several SILs) will be required to get the air vehicle tested, troubleshooting and testing autonomous decision-making will require specialized capabilities.

Provided that the system architecture supports it, it may be possible to isolate the decision-making hardware and software, feeding them synthetic external sensor, communications, and navigation inputs and conducting tests in a virtual environment. This would be valuable, particularly for the initial debugging and trouble-shooting part of the development process. It would be particularly valuable if there is a capability to perform test runs in accelerated time. Real-time runs are possible, but force two or three-shift continual operation of the SIL.

That "fast forward" capability will be needed. Because once the decision-making functions are working reliably, the test team should be presenting them with a wide range of situations and observing what the responses are.

Here, the comparison with automobile navigation systems is particularly valuable. Given a specific starting situation, a specific desired end-state, and priorities for distance or time, a highly autonomous system should produce an answer – that the test team can predict, because they know what those inputs were. If the results are not as predicted, the database should be examined to ensure that the start-state, end-state, and other applicable data are correct. A car navigation system may well take the driver down a 50 kph two-lane road in preference to an 80 kph highway if the database shows the two-lane road as having a 100 kph speed limit.

But if the database information is correct, the decision-making logic isn't working properly, and further inquiries with the software team are needed to determine whether there is an error in the logic itself, an error in the coding…or something unanticipated.

An example of the latter was the behavior of the RQ-4 Global Hawk when it was flying out of Eglin AFB, FL, in 2000 for the LINKED SEAS exercise. The aircraft would begin its final approach – then wave off. Commands to repeat the approach produced the same result. The aircraft was ultimately landed safely, but there was much head-scratching. Investigation showed that the waveoffs were due to the early part of the approach being over water, and the rise of the land as the aircraft passed over the shoreline – which was interpreted by the guidance system as an out-of-limits descent rate. It was an unanticipated characteristic that required correction, but not a fault of the waveoff logic. The system had done precisely what it had been programmed to do, but the software team had not anticipated terrain effects on final approach.

Which brings up a significant issue that the test team will face. Most modern aircraft use copious amounts of software – code that is often written by people who may have no idea whatsoever of how an airplane works, or of the environment in which it must operate.

In any event, the autonomy systems must be thoroughly tested in the SIL environment. Not merely tested once or twice, nor placed in only a handful of simulated test situations, but exercised in a wide range of simulated conditions. A good SIL test program will give the flight test team an autonomous decision-making system in which they can have some confidence.

21.7 FLIGHT TESTS OF HIGHLY AUTONOMOUS SYSTEMS

In planning and executing flight tests of highly autonomous aircraft, safety must come first. Unmanned aircraft worry range safety officers and test organization managers under the best of conditions, and the idea of a highly autonomous system going murderously rogue has been a staple of science-fiction stories for the last sixty years. Extreme caution is therefore indicated.

Clearly, the SIL can be used, both as a trainer for the aircrew and engineers, and as a way to validate planned emergency procedures. Test programs for 4th Generation UAVs invested considerable time in SIL training runs, which significantly reduced risk and helped with test productivity.

As mentioned before, an obvious precaution is to conduct envelope expansion and performance testing first using a low level of autonomy, such as autonomous execution of pilot decisions, and leave testing the autonomous decision-making for later.

Risk can also be reduced by using a surrogate air vehicle known to be airworthy as a testbed. This will not always be feasible, but where it is, this can let the decision-making capabilities be wrung out with greater safety. An example of this is the Cooperative Operations in Denied Environments (CODE) experiments, which mated advanced autonomy with an off-the-shelf Tigershark UAV to provide an economical test platform.

Another safety precaution is to have a manual override capability to allow intervention when the highly autonomous decision-making system fails or behaves unpredictably. At CODE tests conducted in 2018, use of this “snatchback” capability was routine – although to be fair, the system being experimented with was not terribly mature. Nevertheless, the ability for the aircrew to take over the decision making is essential for safety, even if it requires the incorporation of special software provisions to permit it.

With the safety considerations addressed, planning can proceed to the actual test events.

These should start with scenarios with predicted outcomes known to the test team. Simple scenarios at first, then build in complexity. If the autonomous decision-making system behaves as predicted, the test team can proceed to add “pop-up” requirements. These might include ad-hoc targets, surface-to-air missile sites, or simply civil air traffic to be avoided.

For systems intended to cooperate with other systems, the numbers of units can be increased…and it may be worthwhile to make some of those units virtual if that is possible. Virtual participants can be deleted quickly and without risk.

Testers should note that arranging a highly complex test environment can become very expensive. While virtual and simulated testing can add confidence, a final test with actual hardware is essential. Test teams should look for opportunities to participate in military exercises when the system maturity will support it, and to pool their test resources to pay for a fully representative test environment to conduct their final assessment of operational utility. Such testing will often highlight significant problems that appeared minor in technical testing – and may also highlight significant strengths that were previously unnoticed.

Finally, highly autonomous systems intended for military applications will require a final assessment of military utility in a representative operational environment. This will definitely demand participation in a military exercise, possibly more than one. Experience with previous unmanned systems has shown that it is prudent to introduce the system into small-scale exercises prior to the big "graduation" exercise. Experience has also shown that initial planning needs to start at least one year ahead of the planned major event…and often more.

It's also worth reminding flight test engineers new to the field to observe and note everything, whether or not it is part of the official test maneuver. Significant issues can arise at any time. This will be particularly true for highly autonomous aircraft. The most serious problems may well manifest themselves in the "routine" parts of the flight.

21.8 POTENTIAL PROBLEM AREAS

Testing is inherently risky. Experience with unmanned aviation teaches that UAVs are particularly vulnerable, more so than manned aircraft of equivalent complexity. Testers of highly autonomous systems will need to exercise wariness.

The first likely problem area is inadequate testing of the basic airframe. The test team may find itself under pressure to rush through testing of the airframe and get on to the much more glamorous autonomy testing. Air vehicle testing is mundane, autonomy is exotic. Until the airframe fails…then the high-visibility autonomous aircraft becomes a high-profile mishap, with all the problems that entails. The test team must maintain strict discipline in the test program, and get the air vehicle working before proceeding to try high levels of autonomy.

The second likely problem area will be pressure to use the highly autonomous capabilities as a safety backup for airframe testing. This sort of thing has happened already in a current test program, which relied on autonomous "fly-home" lost-link response to deal with a command uplink that was known to be unreliable, instead of having a proper backup link. This worked...until it didn't. And the temptation to use highly autonomous capabilities as a backup will be tremendous. But it should be resisted, for that is the path to high-visibility failure.

The third likely problem area will be a rush to get high levels of autonomy into flight test prematurely. Testers must remember that the software to deliver this capability is being written by computer programmers who may not have much aviation experience, and frequently a background in an industry that considers it acceptable to sell late-beta-test software to paying customers. Not to mention the temptation to rush the capability into flight test prematurely. The test team will need to exercise great care to ensure that the software needed to control high levels of autonomy has been properly tested in the SIL prior to flight testing. And then to execute a disciplined, cautious test program.

21.9 SWARMS

The advent of highly-autonomous systems leads to another desirable UAS capability: swarming. Generically speaking, a "swarm" is a collection of UAVs operating cooperatively or collaboratively. This is achievable when system-level autonomy allows a "one-to-many" relationship of operators to aircraft. Robust implementations of this concept allow a single operator to control multiple UAVs as if they were a single entity.

Swarming is typically conducted by small UAVs due to their relative ease of employing in mass quantities. A homogeneous swarm will typically consist of multiple similar UAVs, while a heterogeneous swarm will typically contain multiple types of dissimilar UAVs. A "mothership" with subordinate drone UAVs would be one example of a heterogeneous swarm: a single sophisticated UAV contains datalink connectivity and processing power to guide the actions of subordinate unsophisticated UAVs.

In military operations, there is a variety of use-cases.

- Multiple sensors of a similar type expanding both the field of regard and field of view
- Multiple sensors of a similar type sharing a single field of view from multiple angles
- Multiple sensors of different types collecting diverse data of a single target or region (multiplatform sensor)
- “Daisy-chaining” line-of-sight (LOS) communications links to enable beyond-line-of-sight (BLOS) communications
- Distributed loads for logistics delivery
- System-of-systems redundancy through greater tolerance to individual failures or attrition

When testing a UAS swarm, it is useful to treat every individual UAV as a subsystem and apply a “system-of-systems” approach. Strictly define each subsystem’s intended function, performance parameters, and measures of effectiveness. If possible, test individual subsystems in isolation prior to aggregating into a swarm. Once individual subsystems’ behavior is fully categorized, their contributions to the swarm-level system performance can be better understood.

21.10 CONCLUSION

Highly autonomous aircraft, capable not merely of executing decisions made by the aircrew but of making many of those decisions for themselves, are rapidly entering flight test. Experience with the last generation of unmanned aircraft teaches us that early planning is essential for safe and productive test execution with these advanced unmanned platforms.