A Briefing on Metrics and Risks for Autonomous Decision-Making in Aerospace Applications

Susan A. Frost[1]
NASA Ames Research Center, Moffett Field, CA, 94035

Kai Goebel[2]
NASA Ames Research Center, Moffett Field, CA, 94035

and

Jose Celaya[3]
SGT, Moffett Field, CA, 94035

Significant technology advances will enable future aerospace systems to safely and reliably make decisions autonomously, or without human interaction. The decision-making may result in actions that enable an aircraft or spacecraft in an off-nominal state or with slightly degraded components to achieve mission performance and safety goals while reducing or avoiding damage to the aircraft or spacecraft. Some key technology enablers for autonomous decision-making include: a continuous state awareness through the maturation of the prognostics health management field, novel sensor development, and the considerable gains made in computation power and data processing bandwidth versus system size. Sophisticated algorithms and physics based models coupled with these technological advances allow reliable assessment of a system, subsystem, or components. Decisions that balance mission objectives and constraints with remaining useful life predictions can be made autonomously to maintain safety requirements, optimal performance, and ensure mission objectives. This autonomous approach to decision-making will come with new risks and benefits, some of which will be examined in this paper. To start, an account of previous work to categorize or quantify autonomy in aerospace systems will be presented. In addition, a survey of perceived risks in autonomous decision-making in the context of piloted aircraft and remotely piloted or completely autonomous unmanned autonomous systems (UAS) will be presented based on interviews that were conducted with individuals from industry, academia, and government.

I.   Introduction

The Aviation Safety Program (AvSafe) under the Aeronautics Research Mission Directorate (ARMD) at NASA is investigating the safety aspects that are necessary to realize the full potential of the Next Generation Aircraft Transportation System (NextGen). State-of-the-art networking and data acquisition and data fusion technology will continually update NextGen's state data and share relevant information with pilots and air traffic controllers. Aircraft will be able to rapidly and safely adjust to changing factors such as weather, traffic congestion, flight trajectories, aircraft spacing, and any terrestrial or airborne security concerns. Working with partners, AvSafe provides fundamental research on existing safety challenges and on new and emerging challenges created by the transition to NextGen, including significant increases in air traffic, introduction of new vehicle concepts, continued operation of legacy vehicles, increased reliance on automation, and increased operating complexity. Under the umbrella of NASA's Aviation Safety Program, the System-wide Safety Assurance Technologies (SSAT) Project

---

[1] Computer Engineer, Intelligent Systems Division, PO Box 1, M/S 269-3, susan.frost@nasa.gov.
[2] Senior Scientist, Intelligent Systems Division, PO Box 1, M/S 269-3, kai.goebel@nasa.gov.
[3] Senior Engineer, Intelligent Systems Division, PO Box 1, M/S 269-3, jose.r.celaya@nasa.gov.

provides knowledge, concepts and methods to proactively manage increasing complexity in the design and operation of vehicles and air transportation systems, including automation design tools and prognostic algorithm design.[1]

Significant advances in several technology areas will enable future aeronautic systems to safely and reliably make decisions autonomously, or without interacting with humans or other technological systems. By using system health monitoring information to determine remaining useful life (RUL) of components of an aircraft, autonomous decision-making algorithms may be able to mitigate the detrimental effect of degrade health and increase the safety of aircraft systems operation. Such autonomous decision-making could enable aircraft to achieve mission performance and safety goals despite the presence of an off-nominal state, or with slightly degraded components. Some key technology enablers for this type of autonomous decision-making include: the maturation of the prognostics health monitoring (PHM) field, novel sensor development, and the considerable gains made in computation power versus system size and data processing bandwidth.

The science of prognostics involves modeling the underlying physics of a component or subsystem, modeling the physics of damage propagation mechanisms in the subsystem, determining criteria for an end-of-life threshold, and developing algorithms to propagate damage into the future that are robust to uncertainty. Sophisticated algorithms and physics based models coupled with these technological advances, enable prognostic health monitoring systems to reliably predict remaining useful life of components or subsystems. Decisions that balance mission objectives and constraints with remaining useful life predictions can be made autonomously to maintain safety requirements and achieve mission performance. This autonomous approach to decision-making will come with risks and benefits to the aerospace community.

In this paper, we first provide a survey of previous work to quantify and assess autonomy in aerospace systems. A applicable to piloted, remotely piloted, or unpiloted aircraft. Analysis and assessment of risks in autonomous decision-making with input from government, industry, and academic stakeholders will be presented.

II. Autonomous Systems

Humans design and build many types of complex, dynamical systems that operate in different time-scales in unpredictable, uncertain, and sometimes hostile environments. Often, some degree of autonomy is designed into systems and subsystems to address a number of issues, including, but not limited to: safety, cost, reliability, efficiency, robustness, accessibility, and physical size. The decision to include autonomy in a system is affected by many factors, such as availability, expense, and reliability of appropriate technology, the ability to verify and validate the autonomy for safety, and public trust of the autonomous system. Usually complex systems have some level of human interaction during operation, and therefore are not totally autonomous, with exceptions including robotic missions to distant locations in space where time and distance inhibit communication for operations. In this paper, *autonomous systems* include systems with some form of human interaction.

Autonomy in aerospace systems has been mostly limited to uninhabited aerial vehicles or systems (UAV, UAS) and planetary flight vehicles. The American Institute of Aeronautics and Astronautics defines autonomy for UAVs as "a system's own ability of sensing, perceiving, analyzing, communicating, planning, decision-making, and acting, to achieve its goal as assigned by human operators."[2] It is important to distinguish between the words *autonomous* and *automatic*. In Bruce Clough's paper on the subject of autonomy for UAV's, he compares the functions of an automatic system with an autonomous system, describing an automatic system, such as an autopilot, as one that executes plans exactly as they are programmed, whereas an autonomous system, such as an autonomous guidance system, can decide to execute actions without outside influence.[3] While this paper here is focused on autonomy, and not automatic systems, we consider the decision to use automation a subset of autonomy.

A. Levels of autonomy

Much of the early work modeling the interaction between humans and computer systems is documented in the seminal book *Telerobotics, Automation, and Human Supervisory Control* by Thomas Sheridan[4]. Several important insights and models are developed in the book although the examples are mainly telerobotic systems (where the human is issuing commands to the system remotely) or human supervisory systems. In particular, Sheridan provides insight into the effects an operator's trust, or lack of trust, has on an autonomous system. Trust in autonomy is the perceived technical and operational reliability and the confidence placed in the autonomy by a human.

Sheridan's work was later extended to create a framework to assist in the design of autonomous systems with human interaction[5]. Drawing from the wealth of literature on human factors studies, i.e., studies investigating human interaction with computerized systems, the authors identify what they consider to be the most relevant issues affecting autonomous systems with human interaction. A model for types and levels of autonomy in systems with

human interaction is proposed along with criteria to determine the type and level of autonomy to design into a system with human interaction.

The authors use a four-stage model of human information processing as a basis for formulating the types of autonomous functions a system can perform. The four types proposed for the framework include: information acquisition, information analysis, decision and action selection, and action implementation. The relationships between the types as these authors saw them are summarized in fig. 1. More recent studies of methods for determining the level of autonomy to design into a human spaceflight vehicle have adopted a similar decomposition of the types of autonomous functions for a spaceflight vehicle[6].
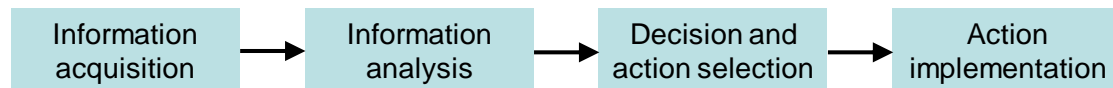
Information acquisition → Information analysis → Decision and action selection → Action implementation

Fig. 1. Categories of autonomous functions.

A measure of the level of autonomy of a system is desirable for system design and evaluation. A level of autonomy (LoA) value can be assigned for each type of autonomous function. The LoA is a number scale that represents low to high levels of automation. The lowest level represents no autonomy (i.e., total human control) and the highest level represents total autonomy with no human interaction. The ten levels of autonomy for decision and action selection, based on a taxonomy originally proposed by Sheridan are reproduced in Table 1 (using consistent grammar). Levels for the other types of autonomy are not presented in the study since the authors expected technology to evolve in these areas and there was a lack of empirical research on the interaction between those types of autonomy and human performance.

Table 1. Levels of autonomy of decision and action selection.

Level	Value	Computer or automation system action
HIGH	**10**	**Decides everything, acts autonomously, ignoring the human**
	9	**Informs the human only if it (the computer) decides to**
	8	**Informs the human only if asked**
	7	**Executes automatically, then necessarily informs the human**
	6	**Allows the human a restricted time to veto before automatic execution**
	5	**Executes a suggested action if the human approves**
	4	**Suggests one alternative**
	3	**Narrows the selection down to a few**
	2	**Offers a complete set of decision/action alternatives,**
LOW	**1**	**Offers no assistance; no autonomy, human makes decisions and takes actions**

The proposed model is intended to provide a framework for applying criteria to determine the type and level of autonomy to design into a system with human interaction. Design of an autonomous system is performed by decomposing the system into the framework described above and applying primary and secondary evaluation criteria. The consequence of autonomy on a human operator's performance is the primary evaluation criteria applied in this framework. Based on documented studies of types of human performance affected by autonomy, the authors chose the following four categories for evaluating the consequences of autonomy on human performance:
1) mental workload
2) situation awareness
3) complacency
4) skill degradation

Mental workload refers to the mental efforts a human must make to perform tasks or achieve goals within a temporal setting. Generally accepted components of mental workload include sensory input, cognitive processing, and psychomotor response. Situation awareness refers to the human's awareness of the states of a system and its operating environment. A human interacting with an autonomous system can become complacent about the process

or task that is being performed, trusting the computer to correctly achieve all of its goals or objective and possibly to even respond to off-nominal situations. The human may go so far as to cease attending to stimuli that would indicate a problem with the system being monitored. Skill degradation can occur when the human operator rarely performs a skill due to the fact that the autonomous system is always performing the tasks that require those skills.

The secondary design criterion for the proposed framework balances reliability and costs of autonomy. Both the costs of decisions and the costs of action outcomes are included in this criterion. Design would then involve applying the two design criteria to determine the optimal level of autonomy to include in a system. Some limitations of the framework noted by the authors are the absence of suitable rules or guidance on selection of levels of autonomy and on assignment of weights for the criteria.

The Observe, Orient, Decide, and Act (OODA) loop[8], developed by John Boyd and presented in his presentation "The Essence of Winning and Losing," was designed to describe the battlefield decision-making process required to win against any opponent, see fig. 2. The OODA loop has been used for successful design and analysis of autonomous systems[6]. The basic premise of the OODA loop is to cycle through the loop, using information, decisions, or actions to inform the other elements of the loop. By understanding the opponent's decision-making process and thereby anticipating his or her decisions, hopefully the actor can cycle through the loop before the opponent can make tactical changes. The task decomposition is similar to the decomposition shown in fig. 1; however, there is an idea of feedback between some of the components.
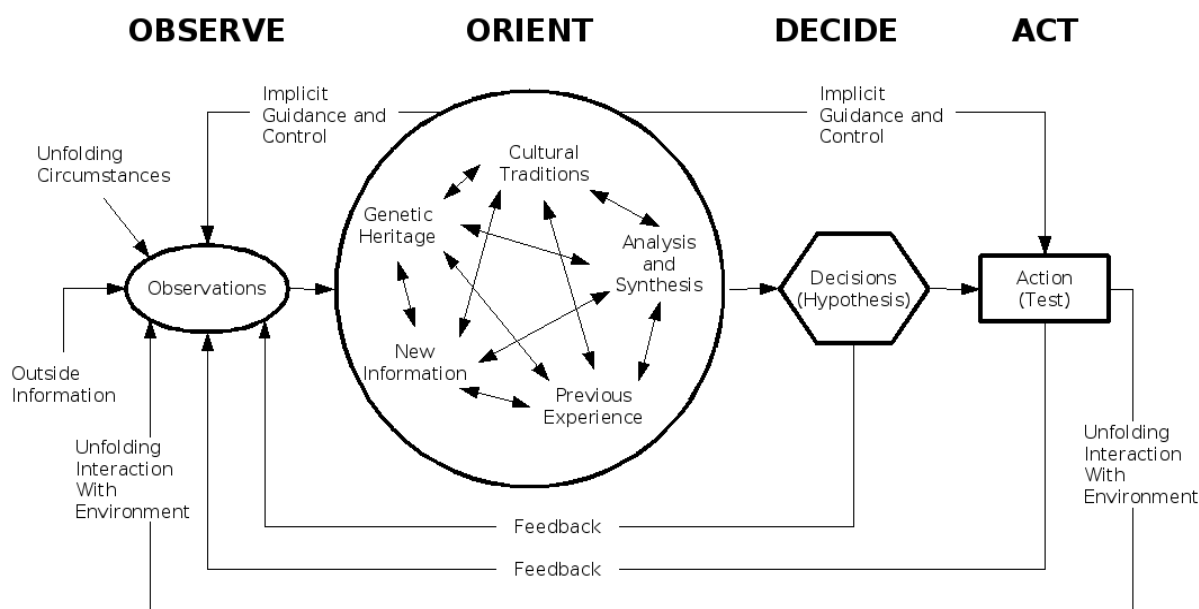

Fig. 2. John Boyd's OODA Loop Diagram[8]

Observations are made using information from the environment, decisions or hypotheses formed from a previous cycle, and observations of outcomes of actions performed in a previous cycle. During the orient task, information from the observe task is analyzed and synthesized in an attempt to understand the opponent and his or her next move. The observations are analyzed, filtered, and synthesized using a number of contexts, such as genetic heritage, cultural traditions, new information, and previous experiences. The decide task takes the synthesized information and makes decisions or forms new hypotheses which are then passed to the act task and back to the observe task. In this way, a hypothesis can inform the observe task about relevant observations to be made. The act task may perform an action or it may perform no action at all. The action or lack of action is fed back to the observe loop, which then can observe the unfolding interaction of the action with the environment. The OODA loop is really many loops working together. By observing the environmental changes resulting from action or inaction and by making observations that are informed by a hypothesis, the orient and decision-making tasks can improve their effectiveness.

B. Decomposition of autonomous systems

The Charles Stark Draper Laboratory proposed a hierarchical decomposition of autonomy requirements in the context of Unmanned Combat Air Vehicles (UCAV)[9]. This effort involved developing autonomy requirements for four mission scenarios, including: air-to-ground missions, close air support, air-to-air missions, and logistics re-supply. Information from internal and external mission experts was distilled to understand the current practice using manned aircraft for these missions and how the missions might be accomplished using UCAVs. The hierarchical approach decomposes the missions into components that were shared among the missions and those that were unique to a mission. Five mission phases were identified as being shared by the four mission scenarios, including:

    I.   Pre-mission plan
   II.   Launch
  III.   Enroute operations
  IV.   Engagement or delivery
   V.   Land and secure

To further decompose the missions, ten elements were identified that apply throughout each mission. Among the ten elements, the following are common to non-military missions:

1) Intelligent autonomy
2) Sensing for situational awareness
3) Communication
4) Vehicle health management and monitoring
5) Collaboration
6) Data security
7) Re-evaluation of mission feasibility

The mission phases were further sub-divided and details were provided in three areas that were expected to illuminate potential difficulties for autonomous aircraft, including:

a) Coordination between UCAVs and manned assets (fighter aircraft, carrier ATC and crew)
b) Contingencies, or conditions or situations that might arise to affect the mission
c) Situational awareness requirements

Using the hierarchical decomposition of the missions described above, requirements for UCAV missions were derived. Since the intent of that work was to examine autonomy challenges driving design of future autonomous platforms, attention was focused on missions that were considered more challenging for UCAVs.

A team of specialists for the National Academy of Sciences reviewed the Draper Lab research effort[42] presented in ref. 9. While ONR's vision of *complete* intelligent autonomy as the goal of all future UAV/UCAV systems was ruled as unrealistic, the review team valued the insights into future research and development needs gained from the detailed mission analyses. Software development, complexity, and cost were seen as critical elements affecting autonomous systems. The reviewers cited significant technology risks due to the lack of adequate tools providing a systematic approach for partitioning functions between machines and humans and for assessing benefits of autonomy. Human trust of an autonomous component or system was deemed to be the most difficult, yet important, attribute to attain when deploying autonomous systems.

A later report describing robust autonomy and motivating scenarios defines a *robust* autonomous system to be an autonomous system that is flexible and adaptable, capable of operating successfully in a dynamic, uncertain and possibly hostile environment, and in the face of imperfect information about both itself and its environment[10]. An approach combining functional and hierarchical decomposition with closed-loop planning and control is proposed to achieve robust autonomy. Problem decomposition is used to make the problem more tractable. First, the problem is decomposed into the following functions: perceiving, reasoning, and acting. Then the problem is decomposed into a hierarchy of smaller, partially coupled problems that can be solved independently, for fixed coupling constraints. Top levels of the hierarchy deal with the overall problem goals and constraints and have longer time horizons and lower levels of detail. Bottom levels of the hierarchy deal with detailed goals and constraints and have shorter time horizons and higher levels of detail. The researchers at Draper Labs had some success with their approach for specific applications; however, they acknowledged that control of a hierarchical planner is a difficult task, especially in dynamic and uncertain environments. They are continuing to develop a meta-controller to control the flow of data and commands within the hierarchy to help achieve overall consistent optimized system solutions in uncertain environments.

C. Autonomy metrics in robotic and aerospace systems

When evaluating autonomous systems, it is desirable to have metrics for meaningful comparisons. Assessing the ability of an autonomous system to perform its task is a relevant goal of a metric. The so called *intelligence* of an autonomous system cannot be used to predict its ability to attain goals. Thus, intelligence metrics such as the Turing Test[11] are not considered here. Several research teams have worked on categorizing and defining metrics for autonomous systems that describe the Levels of Autonomy (LOA) of systems[3,9,12-15].

Researchers at Los Alamos National Laboratory in New Mexico developed the "Mobility, Acquisition, and Protection" (MAP) survival metric space[12]. Six levels were used to describe a simple robot's ability to perform the following tasks:

1) Move in its environment
2) Acquire energy
3) Protect itself

A *radar chart* is a simple way to graphically represent multivariate data in two dimensions. Researchers from the Air Force Research Laboratory (AFRL) at Wright-Patterson AFB used the three MAP metrics to form an axis of a radar chart. They showed the versatility of the MAP metric space by demonstrating that disparate objects such as an ant, a toaster, and a human can all be represented on one MAP radar chart. However, they found MAP to be inadequate when applied to assess autonomy of UAVs[3].

Besides evaluating the MAP metric space, AFRL explored work done by Draper Labs (see previous section) and decided to combine elements of each metric with some of their own ideas to develop the Autonomous Control Level (ACL) chart for UAVs[3]. See Table 2 for a copy of the ACL chart that was given in ref. 13. Eleven levels of autonomy were represented on the chart from the lowest level of autonomy to the highest level of autonomy. The lowest level, level 0, corresponds to a remotely piloted vehicle and the highest level, level 10, corresponds to a completely autonomous vehicle. The original metric space had three axes, but it was deemed inadequate to separate important issues, leading to the adoption of four axes based on Boyd's observe, orient, decide, and act loop. AFRL obtained satisfactory results when the ACL chart was applied to assess UAV efforts and to determine future UAV research directions.

Table 2. AFRL Autonomous Control Level (ACL) chart.

Level	Level Descriptor	Observe	Orient	Decide	Act
		Perception / Situational Awareness	Analysis / Coordination	Decision Making	Capability
10	Fully Autonomous	Cognizant of all within Battlespace	Coordinates as Necessary	Capable of total Independence	Requires little guidance of any sort
9	Battlespace Swarm Cognizance	Knows intent of self and others (friendly and threat) in a complex/intense environment; on board tracking	Group Strategic Missions Assigned; Threat Tactics Inferred	Distributed Tactical Group Planning; Individual mission decisionmaking; chooses targets	Group executes mission with minimal supervisory assistance
8	Battlespace Single Cognizance	Proximity Inference - intent of self and others (friendly and threat); Reduced dependence on off-board data	Strategic Group Goals Assigned; Threat Tactics Inferred; Aided Target Recognition	Coordinated Tactical Group Planning; Individual Task Planning and Execution; Chooses targets of Opportunity	Group executes mission with minimal supervisory assistance
7	Battlespace Knowledge	Short track awareness - History and predictive battlespace data in limited range, timeframe, and numbers; Limited inference supplemented by off-board data	Tactical group goals assigned; Enemy location estimated	Individual task planning/execution to meet goals	Group accomplishment of tactical goal with minimal supervisory assistance
6	Real Time Multi Vehicle Cooperation	Ranged awareness - on board sensing for long range, supplemented by off-board data	All below plus Enemy location sensed/estimated	Coordinated trajectory planning and execution to meet goals - group optimization	Group accomplishment of tactical goal with minimal supervisory assistance Possible close air space separation (1-100 yards.)
5	Real Time Multi Vehicle Coordination	Sensed awareness - Local sensors to detect external targets (friendly and threat) fused with off-board data	All below with Prognostic Health Mgmt; Group diagnosis and resource management	On-board trajectory replanning - optimizes for current and predictive conditions; Collision avoidance	Group accomplishment of tactical plan as externally assigned; Air collision avoidance; Possible close air space separation (1-100 yards); Formation in non-threat conditions
4	Fault/Event Adaptive Vehicle	Off-board Awareness - friendly systems communicate data	All below plus ROE assigned; Inner loop changes reflected in outer loop performance	On board trajectory replanning - event driven; Self resource management; Deconfliction	Self Accomplishment of tactical plan as externally assigned
3	Robust Response to Real Time Faults/Events	Health/status history and models	Tactical Plan assigned; RT health Diag; Compensate for most control failures and flight conditions (i.e. adaptive inner loop control)	Evaluate status vs. required mission capabilities; Abort/return to base if insufficient	Self Accomplishment of tactical plan as externally assigned
2	Changeable Mission	Health/status sensors	RT Health diagnosis (Does UAV have problem(s)?; Off-board replan (as required)	Execute preprogrammed or uploaded plans in response to mission and health conditions	Self Accomplishment of tactical plan as externally assigned
1	Execute Preplanned Mission	Preloaded mission data; Flight Control and Navigation Sensing	Pre / Post Flight BIT; Report Status	Preprogrammed mission and abort plans	Wide airspace separation requirements (kilometers)
0	Remotely Piloted Vehicle	Flight Control (attitude, rates) sensing; On Board Camera	Telemetered data; Remote pilot commands	None. Off-board Pilot.	Control by Remote Pilot

The Army's System Simulation Development Directorate (SSDD) evaluated the AFRL ACL metrics and found them useful for measuring autonomy of relatively large UAVs operating in the Air Force mission space[13]. The SSDD analysis team modified the ACL chart to tailor its ability to analyze UAVs in their application area, e.g., tactical UAVs operating in an urban environment. Additional satisfaction criteria were included for the ACL levels along each of the four OODA dimensions. The satisfaction criteria were in the form of questions specific to a particular level of autonomy and task.

An ad hoc working group of government and industry experts was developing a framework for autonomy levels of unmanned systems (ALFUS) in the early 2000's[14-15]. Using a three axis system, the ALFUS Detailed Model was created to characterize autonomy levels for unmanned systems and their missions. The axes for the ALFUS Detailed Model are:

1) Mission complexity
2) Human interface
3) Environmental difficulty

To use the framework, the user decomposes a UAV mission into subtasks, and assigns a metric score to each of the three axes for that subtask. Moving from the lowest level subtasks to the highest, a metric score and weight is assigned to each axis for a given subtask. The metric scores and weights are averaged for each subtask level and combined with the next higher level's subtask score. This process continues until the highest level subtask is included, resulting in a composite score for the mission and vehicle. The group reported that guidelines for decomposing tasks, prioritizing and weighting the metrics, correlating interdependencies among metrics, developing scoring scales for the metrics, and integrating the metrics to develop a concise set of indices for the autonomy levels need to be developed[15]. The ALFUS group was absorbed by the SAE-AS4D Committee on Performance Measures.

In a study to develop a method for determining the level of autonomy to design into a human spaceflight vehicle management system, tasks were decomposed into a four-tiered architecture based on the OODA Loop[6]. The authors of this study wanted to devise a questionnaire and scale to assess costs, benefits, and the level of trust a human has in the autonomy. The level of human trust in an autonomous system is a subjective measure; hence the authors looked to the Cooper-Harper scale and questionnaire that is used to rate test-pilot assessments of aircraft handling qualities[16]. A questionnaire and scale was created, combining two metrics in this study, one to assess how much a human user trusts the autonomy in a system and a second to measure the cost/benefit ratio of the system. The objective was for the metric to distinguish systems with undesirable qualities. In particular, the metric was designed to identify systems in which there is 1) a high level of trust in the autonomy and a high cost/benefit ratio of the autonomy or 2) a low level of trust and a low cost/benefit ratio of the autonomy.

D. Analysis of robotic and aerospace systems

Work done to define metrics for robotic systems can be useful when defining metrics for autonomous aerospace systems. In ref. 17, metrics are proposed for several general tasks that robots would perform. Metrics are defined to evaluate the effectiveness, efficiency (in terms of time and effort), ability of a system to present information to a human, confidence or trust generated in a human, and various errors specific to the task. The authors also attempt to account for biasing effects, such as system latency differences, which could impact the metric across different solutions. The authors further suggest giving the system the ability to choose its level of autonomy to improve performance as measured by the metrics. Several papers report on various performance metrics for evaluation of autonomous systems[18-20].

A detailed study of the impact of autonomous system technology on uninhabited aerial vehicles was performed from a systems analysis point of view[21]. A number of different measures and metrics are proposed to estimate costs and benefits of autonomy, to estimate and track mission success, to assess autonomy *intelligence* and *elegance*. According to the authors, *intelligence* measures how well actions, tasks or roles are performed under varying degrees of task and environmental complexity and other associated constraints and conditions. Autonomy *elegance* is defined to be the computation efficiency by which the autonomous vehicle intelligence is implemented. Hence, two autonomous systems can be at nominally equivalent autonomy levels by exhibiting radically different levels of intelligence. First order functional relationships were proposed that relate autonomy metrics to vehicle characteristics and mission requirements. Analysis was performed using the proposed tools and metrics to gain insight into the amount of autonomy to include in a UAV mission. Development and implementation costs were kept to a minimum and vehicle performance was optimized in the context of vehicle autonomy and intelligence, while allowing for acceptable levels of success and risk. The general return on investment trends as a function of intelligence of the autonomy suggests the most benefit when incorporating a mid-level of intelligence for a mid-level of autonomy. The contribution of individual autonomous system technologies were related conceptually and quantitatively to high-level programmatic goals, mission cost, risk, and return on investment criteria. The

quantitative analysis of the relative importance of individual autonomous system technologies identified technologies having statistically significant importance.

An autonomy focused assessment methodology was developed and analyzed in ref. 22. In particular, the methodology is evaluated to consider its applicability and fitness in understanding and demonstrating the perceived strengths and weaknesses of autonomous solution options for representative military requirements. The intent of the assessment tool is to map solution options against requirements and to identify important operational analysis (OA) issues with capability development, procurement, cost of ownership, use, maintenance, and delivery risk. The assessment approach adopted in this study was to quantitatively link key measures such as cost and risk to levels of autonomy. A limitation identified by the study[22] was the inability of the tool to capture the attributes of robustness of the system, that is, the ability of a system to continue working towards a specified goal notwithstanding a number of unpredicted events and degraded sub-system performance. It was deemed imperative that the assessment criteria be able to discriminate robust solution options. Consequently, a range of potential failure events were folded into the criteria. The assessment tool was developed to allow the use of different weighting sets and mappings of the criteria and to apply bounds to the criteria. The study proposes and demonstrates an integrated modeling and simulation approach that is then incorporated within an OA framework for analysis.

The proposed integration of UAS into the civil airspace is prompting agencies throughout the world to examine the implications of autonomous systems. Researchers are developing architectures for system health management on civil UAS to ensure the safe operation of UAS in the civil airspace[23]. A framework for testing robust autonomy of UAS during design and certification has been proposed for implementation in Australia[20].

E. Human trust in autonomous systems

It is important to be able to estimate the level of human trust in an autonomous system, since this affects adoption of the technology by operators. Trust in autonomy is generally needed if the human can choose whether to make use of the autonomy. However, it can be dangerous if humans trust an autonomous system that is unreliable. In ref. 4, the author describes the following seven attributes of autonomous systems that can positively or negatively affect human trust, see Table 3.

Table 3. Seven attributes of autonomous systems that can affect trust[4].

Attribute	Example
Reliability	**System consistently and repeatedly performs or functions correctly**
Robustness	**System can perform under a variety of conditions or circumstances**
Familiarity	**System employs procedures, terms, and cultural norms that are familiar or seem natural**
Understandability	**Human can form a mental model and predict future system behavior**
Explication of intention	**System explicitly communicates how it will behave (like having a contract)**
Usefulness	**System responds or behaves in a way that is useful to the human**
Dependence	**Dependence of the trusting person on the system**

The Air Force Research Laboratory (AFRL) has conducted human factors research on supervision of simulated autonomous UAVs to explore issues associated with automation management[24]. In particular, AFRL has focused on operator performance, situation awareness, and trust in the autonomous system with different levels of autonomy and system reliability. Results of this study indicated that the operators would avoid using the autonomous functionality provided unless they had insufficient time to accomplish the task without assistance from the computer. When the study was recast to reduce the allowed time for operators to accomplish the desired task, the operators rated the work load as being higher and their trust ratings were lower. The study authors concluded that there is a complex relationship between LoA, time limits, and perception of task difficulty and confidence in autonomy. They also suggest exploring additional LoAs that are contingency/task specific and can be changed during a mission.

A recent piloted simulation study of an emergency landing planner (ELP) provided interesting insights into human interaction with autonomy[25]. The researchers developed a sophisticated planner, designed to assist pilots in choosing the best emergency landing site when damage or failures occur in an aircraft. The planner was integrated into the cockpit of a 6 degree-of-freedom full-motion simulator for 757/767 category transport aircraft. Experiments were performed to evaluate the software using crews of professional airline pilots. The ELP proved to be helpful to the pilots in poor weather or wind conditions, where the ranking of best runway was more complex. The

experiments provided interesting insights into the decisions pilots made in poor weather conditions. In one scenario, some of the pilot teams chose a longer runway with poor weather conditions that was lower ranked by the ELP over a shorter runway with better conditions. Generally, the pilots who chose the shorter runway that was ranked highest by the ELP were able to safely land the plane. However, in almost every case where the pilots chose the lower ranked but longer runway, the outcome was a fatal accident in simulation.

III. Autonomous Decision-making using PHM Information

We define *autonomous decision-making using PHM information* to be the process that is carried out in reaction to changed health management information or to changed mission goals. The result of the decision-making is a set of appropriate mitigating actions that are meant to resolve the issue stemming from the degraded state of health or changed goals. The selection process involves, or is, a general decision-making process that is either 1) *reactive*, to ensure immediate proper functionality or safety; or 2) *goal seeking*, to re-plan activities or modify system or subsystem operation to mitigate future events that would require reactive decision-making. The determination of which of the activities are carried out depends to a large degree on the prognostic horizon, or, more specifically, the time to expected failure. If, for example, the time-to-failure is very short (say, in the range of milliseconds), one needs to generally react fast using adaptive control mechanisms at the machine controller level. If the prognostic horizon is in the second range, appropriate contingencies may involve control reallocation or reconfiguration, i.e., the use of other components or subsystems. If the prognostic horizon is even longer, one can consider mission re-planning. Finally, if the prognostic horizon is considerably longer, perhaps even extending beyond the duration of the mission at hand, one could integrate the logistics operations and consider various optimal maintenance actions.

It should be noted that these contingency management actions are realized with very different decision-making techniques. If the contingency management is realized via adaptive controls, then the controller needs to be modified to react to changed objectives. This can be accomplished by either modification of the controller gains, modification of the control modifiers, modification of the control schedules, or modification of the control structure. To account for the short time scale, these solutions may have been pre-compiled and it is just a matter of a look-up to execute the right solution.

Work in integrating fault information has been performed for a long time in fault tolerant control[26-27], but also in life-extending controls[28-29]. Generally, life-extending control attempts to modify control logic or control hardware to influence life consuming factors. This approach has been applied to aircraft engines where reduction of creep damage to turbine blades was accomplished by incorporating an engine damage model into the flight mission. Through optimization over flight time, fuel consumption, and component damages overall operation cost was reduced[29]. In other work, a model-based detection scheme was coupled with controller adjustments that were pre-compiled off-line using evolutionary algorithms to identify optimal biases and adaptive control[30].

A PHM-based reconfigurable controls framework that is part of hierarchical control architecture is defined in ref. 31. The low-level reconfigurable controller is defined as a time-varying multi-objective criterion function with appropriate constraints to determine optimal set-point reconfiguration. A set of necessary conditions are established to ensure the stability and boundedness of the composite system. The integration of prognostics in the control loop using receding horizon control principles is presented in ref. 32. Life-extending control of multi-degree of freedom mass-beam structure where the control policy is formulated based on the analysis of stochastic time series data of sensors is described in ref. 33.

For situations with somewhat larger prognostic horizon, re-planning and rescheduling may be the appropriate action (perhaps in conjunction with controller changes). An agent-based architecture that integrates fault detection and re-planning activities is described and demonstrated with fault injection aboard a spacecraft[34]. A similar fault tolerant planning environment is described in ref. 35.

Finally, if the prognostic horizon is sufficiently large, one can consider various long-term actions that might include maintenance and inclusion of logistics elements. Challenges arise from the large amount of different information pieces upon which a decision-maker has to act. Consider, as an example, a decision support system for use in operational decision-making in the context of running missions and performing maintenance. The decision support system enables the user to make optimal decisions based on his/her expression of rigorous trade-offs through guided evaluation of different optimal decision alternatives under operational boundary conditions using user-specific and interactive collaboration. An Evolutionary Multi-Objective Optimization can perform the search in this space and generate a set of optimal solutions (the "Pareto frontier"). This will result in the identification of alternative mission allocations and maintenance plans that are optimal along IVHM-specific objectives (e.g., overall mission success, safety, and maintenance cost)[36]. Because this set of optimal solutions may still be too large to just choose any arbitrary one, it is required to further down-select the solutions using appropriate techniques that allow

an assessment of the trade-offs across numerous dimensions. Depending on present and future requirements, the end-user would use context-dependent criteria to constrain the number of solutions using interactive tools to finally select the operational plan that best meets the field requirements. Such an interactive method would also allow users to employ what-if situations, permitting them to manually test the robustness of the solution. Finally, the mitigation of faults in an autonomous setting through stochastic optimization integrating fault risk assessment and risk management techniques is described in ref. 37.

IV. Perceived Risks in Acceptance of Autonomous Decision-making

Interviews were conducted with a few dozen individuals from industry, government, and academia. The objective of the interviews was to determine perceived risks in acceptance of autonomous decision-making in the context of piloted aircraft. The results are summarized below.

A. Human performance consequences in mixed initiative scenarios

There are many aspects of autonomous decision-making using PHM that could impact the performance of humans interfacing with the aircraft, including pilots, mechanics, air traffic controllers, and ground operations crews. In mixed-initiative scenarios, i.e., where autonomous systems have to interact with humans in some ways, human performance may be impacted in positive ways but also in negative ways. Finding the appropriate balance of autonomy is one key aspect to increasing the positive performance consequences. For mixed-initiative scenarios, a medium level of autonomy appears to be generally the best approach[4,6,10,22,23]. A higher level of autonomy is perceived as having possible negative consequences on human performance. Below is a summary of the types of human performance that can be affected by mixed-initiative systems.

1. User workload

Autonomous decision-making can reduce the pilot workload by performing tasks without pilot interaction. If the actions are completely reliable, the benefits are clear. However, since there are few cases of 100% reliability in complex systems, there is a change of increased workload because the pilot or ground crew would have additional work to clear the erroneous action from the system. Where an autonomous system requires increased level of interaction with a pilot input the benefits may be negated if the user could have come up with the same resolution in the same amount of time. Further, autonomous decision-making could increase user workload do to the complexities of communicating the changes to the operator. If frequent active scans are required to keep abreast of the latest changes, the benefits are again diminished.

2. Situation awareness

If changes in the state or dynamics of a system are masked by autonomous decision-making, the human could encounter a situation where critical decisions are made with incomplete or inaccurate information[15]. This problem would be compounded by time constraints on the human operator. Humans might miss important trends in low criticality areas where the system is autonomously changing due to PHM information without appropriate human notification. In 1991, there was an SAS incident in which a McDonnell Douglas MD-81 aircraft was not properly de-iced[38]. The engines ingested clear ice from the wings, which broke free during take-off, causing engine surges. The surges destroyed the engines, requiring the pilots to crash land. Fortunately, there were no fatalities. About 25 seconds after take-off, the pilots perceived banging, vibrations, and jerks in the aircraft. They responded to an engine 2 surge by throttling back the engine. However, due to the operating point in the flight envelope, the throttle was simultaneously controlled by an automatic thrust restoration (ATR) system, which increases throttle setting with altitude. The ATR system caused additional engine surges that eventually destroyed the engines. The airline was unaware of the existence of the ATR system; consequently they had not trained the pilots on proper response to the situation. This case illustrates automation causing a lack of situation awareness for the crew, since the crew was unaware of the ATR system and its affect on throttle position. It is also noteworthy that the ATR system designers did not appear to include the OODA Loop philosophy of having the automation observe the results of its actions, i.e., that the engines were surging.

3. Complacency of humans

If the autonomous decision-making is highly reliable, but not completely reliable or robust to unmodeled dynamics, then the human operator could relax their vigilance for off-nominal behavior, missing the occasional time that there is an undetected problem. When the human work load is the highest, it is natural for the human to become even more reliant on autonomously controlled tasks. Even though the human might know that a system has a certain

probability of making an error when suggesting options, the human can get distracted by the information presented and the choices to be made, forgetting that either of these two things could be incorrect.

4. *Skill degradation*

With a lack of use of certain skills, humans are liable to forget certain elements. Literature from cognitive psychology states that humans forget information and their skills decay over time with disuse[4]. Studies have shown that when an autonomous system fails, an intermediate level of autonomy tends to result in better performance by the human operators, presumably because they have had more skill practice[4].

5. *Human trust in the system*

Lack of trust in autonomous systems may be some of the biggest hurdles in adopting autonomy in mixed-initiative settings. Numerous experimental results suggest that humans have a difficult time giving up control of a complex task that they can perform proficiently, especially when they have to give control to a complex computer system[24-25]. This has in part to do with awareness of the failings of previous autonomous systems but it may also be due to other psychological elements where the human simply wants to "stay in control". Where operators do relinquish their decision-making authority, it is often only due to time constraints[22-23]. The emotional response to autonomous systems may even be manifested in fear[39]. It is expected that autonomy in mixed-initiative systems will slowly gain acceptance, in part through adoption of automated tasks such as auto-pilot and the automatic control of subsystems (such as jet engines).

B. Autonomy reliability

Many of the interviewed individuals cited reliability of autonomous systems, including the reliability of sensors, communication, and algorithms, as being one of the greatest risks. The reliability of the autonomous system must be end-to-end, i.e., it is not sufficient to reliably detect a fault if the response (maybe just simple logging for maintenance) has lower reliability. The guiding mantra is that an autonomous system must do no harm. Proving that an autonomous system is on average more reliable will be difficult because the avoidance of a situation that would have had negative safety implications is hard to validate. The required degree of reliability will be determined criticality of a task.

C. Autonomy effectiveness and robustness

Adding autonomy to a system greatly increases the amount of verification and validation required and with that the cost for deployment. Adding complexity to a system increases also the probability of errors, possibly negating any beneficial effects. If, after all the additional investment of resources, the resulting system is not more effective or at improving safety than a non-automated system (or less robust), then its use is not justifiable. Obtaining certification from the FAA could be another major stumbling block. The difficulties that commercial operators have in obtaining maintenance credits for aircraft with on-board health and usage monitoring systems will likely be magnified for certification attempts of autonomous subsystems. From a commercial perspective, providing the appropriate business case for autonomous systems will be needed to justify adding autonomous functions.

Deployment of autonomy on civil transport aircraft is dependent on development of new methodologies to certify airborne software and to satisfy FAA airworthiness requirements. Adaptive controllers that use real-time system input/output information and some form of on-line learning algorithm to adaptively set controller gains can be seen as providing a low level of autonomy. Current research and development of adaptive flight control for aircraft operating in off-nominal situations shows promise, however many issues exist for their verification and validation by the U.S. Federal Aviation Administration[40].

D. Costs of Decisions

The costs of the decisions made by an autonomous system must be heavily balanced to increase safety of humans affected by the system and secondarily to decrease damage to a system. Autonomous systems that have to choose solutions in a multi-objective environment are faced with the difficulty of making context-dependent choices. These are hard to encode since the plurality of all possible contexts is difficult to encapsulate. Furthermore, the weights of individual objectives may change dynamically due to unforeseen circumstances which further complicate the decision choice.

E. Public perception

In the end, commercial operators need to be sensitive to the way their system is perceived by the public. A bias exists that a human would handle certain situations in a more sensible way than a completely autonomous system[41].

In a hypothetical scenario where a piloted aircraft and an unpiloted aircraft both crash into a school, it would be more likely that public would perceive the pilot as having tried to avoid hitting the school, whereas the autonomous system would be perceived as "cold-hearted." Public belief tends to be that designers are incapable of adding a "humane" aspect to the decision-making used in autonomous systems. The public's perception of autonomy can have serious consequences for airline carriers. Similar issues have been considered for advanced terrestrial applications, such as platooning (autonomous following of several cars on highways)[42]. Currently, there is a slow migration towards the public acceptance of autonomy, for example in parallel parking assist[43] and autonomous distance keeping for luxury cars. However, it is still a long way from adoption in aerospace settings.

V. Conclusions

Prior research into metrics and design for autonomy were presented. At this time, the prospect of adding significant autonomous decision-making on a piloted aircraft is viewed with some degree of concern for the ability of the system to add value without adding risk.

Acknowledgments

References

[1]"Aviation Safety Program Website," National Aeronautics and Space Administration, April 2012, URL: http://www.aeronautics.nasa.gov/programs_avsafe.htm

[2]AIAA Standards Series, "Recommended Practice: Terminology for Unmanned Aerial Vehicles and Remotely Operated Aircraft," Published by AIAA, R-103-2004e, 2004.

[3]Clough, B., "Metrics, Schmetrics! How the Heck Do You Determine a UAV's Autonomy Anyway?" *Proceedings of the 2002 Performance Metrics for Intelligent Systems Workshop (PerMIS-2002)*, 13-15 Aug 2002, Gaithersburg, MD.

[4]Sheridan, T. B., Telerobotics, Automation, and Human Supervisory Control, The MIT Press. 1992.

[5]Parasuraman, R, Sheridan, T.B., and Wickens, C.D., "A model for types and levels of human interaction with automation," *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 30, no. 3, pp. 286-297, May 2000.

[6]Proud, R.W., Hart, J.J., and Mrozinski, R.B., "Methods for determining the level of autonomy to design into a human spaceflight vehicle: A function specific approach," *Proceedings of the 2003 Performance Metrics for Intelligent Systems Workshop (PerMIS-2002)*, 16-18 Sep 2003, Gaithersburg, MD.

[7]Sheridan, T.B., and Verplank, W.L., "Human and computer control of undersea teleoperators," MIT Man-Machine Systems Laboratory, Cambridge, MA, Tech. Rep., 1978.

[8]Boyd, J. R. "The Essence of Winning and Losing," Excerpts in presentation format dated August 2010, April 2011, URL: www.dnipogo.org.

[9]Cleary, M., Abramson, M., Adams, M.B., and Kolitz, S., "Metrics for embedded collaborative intelligent systems," *Proceedings of the 2000 Performance Metrics for Intelligent Systems Workshop (PerMIS-2000)*, 14-16 Aug, Gaithersburg, MD, Draper Tech. Rep. P-3833.

[10]Abramson, M., Cleary, M.., and Kolitz, S., "Steps toward achieving robust autonomy," *Presented at the AAAI Spring Symposium on Robust Autonomy* (but not in proceedings), Mar 2001, Draper Tech. Rep. P-3876.

[11]Turing, A.M., "Computing Machinery and Intelligence," *Journal of Mind Association – A Quarterly Review of Psychology and Philosophy*, 1950, vol. LIX, No 236, pp. 433-60, Oxford University Press.

[12]Hasslacher, B., Tilden, M.W., "Living Machines", Tech. Rep. Los Alamos National Laboratory, 1995.

[13] Sholes, E., "Evolution of a UAV autonomy classification taxonomy," IEEE Aerospace Conference, Big Sky, MT, 2007.

[14]"Autonomy Levels for Unmanned Systems Framework, Volume 1: Terminology, Version 1.1," Huang, H. Ed., NIST Special Publication 1011, National Institute of Standards and Technology, Gaithersburg, MD, September 2004.

[15]Huang, H.-M., Pavek, K. Albus, J., Messina, E., "Autonomy Levels for Unmanned Systems (ALFUS) Framework: An Update," *Proceedings SPIE Defense and Security Symposium*, 2005, Orlando, FL.

[16]Cooper, G.E. and Harper, Jr., R.P., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," NASA TN D-5153, April 1969.

[17]Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., Goodrich, M., "Common metrics for human-robot interaction," *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, paper no. 1-59593-294-1, Salt Lake City, Utah, 2006.

[18]Lewis, A.S. and Weiss, L.G., "Intelligent Autonomy and Performance Measures for Coordinated Unmanned Vehicles," *Proceedings of the Performance Metrics for Intelligent Systems Workshop (PerMIS-2004)*, National Institute of Standards and Technology (NIST), Gaithersburg, MD on August 24-26, 2004.

[19]Crandal, J.W. and Cummings, M.L., "Developing performance metrics for the supervisory control of multiple robots," *Proceedings of ACM Conference on Human Robot Interaction 2007,* Mar. 2007, Arlington, VA.

[20]Perez, T., Donaire, A., Delamberterie, P. and Williams, B., "A framework for testing robust autonomy of UAS during design and certification," *Proceedings AIAA Infotech Conference 2011,* St. Louis, MO, Apr. 2011.

[21]Young, L.A, Yetter, J.A., and Guynn, M.D., "Systems Analysis Applied to Autonomy: Application to High-Altitude Long-Endurance Remotely Operated Aircraft," *Proceedings AIAA Infotech@Aerospace Conference*, Sept 2005.

[22]Heley, D.A., "Autonomous System and Military Challenges - How Might We Assess Applicability?" *Proceedings SEAS DTC Technical Conference*, Edinburgh, Scotland 2007.

[23]Downes, C. G., "ASTRAEA T7-PHM & Contingency Management for Civil UAVs: A Reliability Centered Systems Engineered Approach", *Proc. of the 17th AR2TS- Advances in Risk and Reliability Technology Symposium*, 17-19th April 2007, Loughborough University, Loughborough, UK, pp. 13- 25, ISBN 0-904947-62-9 (2007).

[24]Ruff, H.A., Calhoun, G.L., Draper, M.H., Fontejon, J.V., and Guilfoos, B.J., "Exploring automation issues in supervisory control of multiple UAVs," *Proceeding of the Human Performance, Situation Awareness, and Automation Technology Conference*, March, 2004, pp. 218-222. (Clearance ASC:04-0411, February 2004).

[25]Meuleau, N., Neukom, C., Plaunt, C., Smith, D.E., and Smith, T., "The emergency landing planner experiment," ICAPS-11 Scheduling and Planning Applications Workshop (SPARK), Freiburg, Germany, June 2011.

[26]Litt J.S., Simon D.L., Garg S. and et al., "A survey of intelligent control and health management technologies for aircraft propulsion systems," *Journal of Aerospace Computing, Information, and Communication*, 2004, vol. 1, no. 12, pp. 543-563.

[27]Zhang, Y. and Jiang, J., "Bibliographical review on reconfigurable fault-tolerant control systems," *Annual Reviews in Control* (Elsevier) vol. 32 (2008) pp. 229–252.

[28]Lorenzo, C.F. and Merrill, W.C. , "Life extending control - A concept paper," *American Control Conference, 1991*, pp.1081-1095, 26-28 June 1991.

[29]Guo, T.-H., "A roadmap for aircraft engine life extending control (I)," *Proceedings of the American Control Conference,* v. 5, IEEE, Piscataway, NJ, 2001, pp. 3702-3705.

[30]Rausch, R., Goebel, K., Eklund, N. and Brunell, B., "Integrated fault detection and accommodation: A model-based study," *Proceedings of the ASME Turbo Expo 2005*, v. 1, pp. 561-569, 2005.

[31]Brown, D., Vachtsevanos, G. and Goebel, K., "A prognostic health management based framework for fault-tolerant control," *Proceedings of Annual Conference of the PHM Society 2011*, 2011.

[32]Tang, L., Kacprzynski, G., Orchard, M., Goebel, K., Saxena, A. and Saha, B., "Prognostics-enhanced automated contingency management for advanced autonomous systems," *Proceedings of International Conference on Prognostics and Health Management 2008*, 2008.[33]Khatkhate, A., Gupta, S. Ray, A. and Keller, E., "Life Extending Control of Mechanical Systems using Symbolic Time Series Analysis," *Proceedings of the 2006 American Control Conference,* Minneapolis, MN, pp. 3765 – 3770, 2006.

[34]Rajan, K., Bernard, D., Dorais, G., Gamble, E., Kanefsky, B., Kurien, J., Millar, W., Muscettola, N., Nayak, P., Rouquette, N., Smith, B., Taylor, W., Tung, Y, "Remote Agent: An Autonomous Control System for the New Millennium," *Proceedings of 14th European Conference on Artificial Intelligence*, IOS Press, Amsterdam, pp.726-730, 2000.

[35]Atkins, E., Abdelzaher, T., Shin, K., and Durfee, E., "Planning and Resource Allocation for Hard Real-time, Fault-Tolerant Plan Execution." *Autonomous Agents and Multi-Agent Systems* 4 (1-2): 57-78, 2001.

[36]Iyer, N., Goebel, K. and Bonissone, P., "Framework for Post-Prognostic Decision Support," *Proceedings of 2006 IEEE Aerospace Conference*, 10.1109/AERO.2006.1656108, 2006.

[37]Bole, B., Goebel, K. and Vachtsevanos, G., "A Stochastic Optimization Paradigm for the Development and Validation of Fault Risk Assessment and Risk Management Techniques," *Proceedings of 2012 Infotech@AIAA Conference*, Orange Grove, CA June 2012

[38]Peter Aaap of Sjöland & Thyselius Datakonsulter AB, personal communication, 2011, (Statens haverikommission (SHK) – Sweden, Swedish aviation authority, report number SHK C1993:57).

[39]Bill Mateson, presentation at AIAA Infotech@Aerospace conference in St. Louis, MO on March 29, 2011.

[40]Jacklin, S.A., "Closing the certification gaps in adaptive flight control software," *Proceedings AIAA Guidance, Navigation and Control Conference*, Honolulu, HI, Aug. 2008, AIAA-2008-6988.

[41]Brendan Williams, private conversation, 2011.

[42]Varaiya, P. and Shladover, S., "Sketch of an IVHS systems architecture", *PATH Research Report*, UCB-ITS-PRR-91-3, 1991.

[43]Ingebretsen, M., "In the News, Intelligent Systems," *IEEE*, vol. 25, no. 6, pp 8 – 12, 2010.

[44]"Review of ONR's Uninhabited Combat Air Vehicles Program," Committee for the Review of ONR's Uninhabited Combat Air Vehicles Program, Naval Studies Board, National Research Council, http://www.nap.edu/catalog/9885.html, April 1, 2012.

American Institute of Aeronautics and Astronautics