Measuring the Resilience of Advanced Life Support Systems

Ann Maria Bell
Orbital Sciences Corporation

Richard Dearden
Research Institute for Advanced Computer Science

Julie A. Levri
NASA Ames Research Center

ABSTRACT

Despite the central importance of crew safety in the design and operation of a life support system, the tools commonly used to evaluate alternative Advanced Life Support (ALS) technologies do not currently provide explicit techniques for measuring safety. The resilience of a system, or the system's ability to meet performance requirements and recover from component-level faults, is a fundamentally dynamic property. This paper motivates the use of computer models as a tool to understand and improve system resilience throughout the design process. Extensive simulation of a hybrid computational model of a water revitalization subsystem (WRS) with probabilistic, component-level faults provides data about off-nominal behavior of the system. The data are used to consider alternative measures of resilience as predictors of the system's ability to recover from component-level faults. A novel approach to measuring system resilience using a Markov chain model of performance data is also developed. Results emphasize that resilience depends on the complex interaction of faults, controls, and system dynamics, rather than on simple fault probabilities.

INTRODUCTION

The design and construction of complex systems that are resilient in the face of unexpected events and component failures presents unique engineering and analysis problems. In addition to the complicated, non-linear interactions between the individual hardware elements, the interactions between the system hardware and the control mechanisms also affect the system's ability to recover from component faults. Yet arguably, resilience is the most crucial property of a life support system. This paper examines the critical problem of measuring a system's resilience and proposes some approaches for doing so.

Techniques for measuring resilience allow engineers to directly compare a critical dynamic performance characteristic of two proposed systems. These measures can be used in conjunction with commonly used mass-based measures of the quality of life-support systems to ensure that important system differences in actual performance are not overlooked. These concrete measures also provide a target to optimize against. For example, system resilience can also be the objective function that a learning controller seeks to maximize.

What does it mean for a system to be resilient? There has recently been a great deal of interest in this topic in a wide range of fields, including software engineering, cell biology, sociology and ecology[1]. The discussion in different fields has been loosely organized around the terms 'resilience' and 'robustness', each of which has multiple, overlapping definitions.

The Santa Fe Institute has gathered eighteen definitions of the term 'robustness', ranging from narrow technical definitions to broad ranging qualitative descriptions intended to identify the commonalities across a diverse range of systems [SFI, 2002]. The alternative definitions also help identify "the similarities and differences among terms often used interchangeably with 'robustness' including 'stability,' 'resilience,' 'reliability,' 'persistence, 'survivability,' 'fault-tolerance,' 'plasticity,' etc."

One of the general definitions best suited to resilience in ALS comes from an ecological perspective:*"Robustness is the persistence of specified system features in the face of a specified assembly of insults."* [Allen, 2002] In the case of life support systems, the persistent system features include the continuous supply of breathable air, potable water and edible food to the crew. The assembly of insults consists of unexpected system events such as

[1] Two interdisciplinary groups studying the issues of resilience and robustness maintain websites with overviews, definitions, discussion groups and extensive references. The Santa Fe Institute's Robustness Program seeks to "explore the phenomenology, origins, mechanisms, and consequences of robustness in natural, engineering, and social systems" and can be found at http://discuss.santafe.edu/robustness/
The Resilience Alliance can be found online at http://www.sustainablefutures.net/resilience/resilienceDef.html

unexpectedly large variations in resource usage by the crew and component faults such as tanks leaking, sensors failing, and so on.

Holling distinguishes between 'engineering resilience,' which focuses on "stability near an equilibrium steady-state, where resistance to disturbance and speed of return to the equilibrium are used to measure the property" and 'ecological resilience,' which measures "the magnitude of disturbance that can be absorbed before the system changes its structure by changing the variables and processes that control behavior." Ecological resilience focuses on off-nominal behavior far from equilibrium and the possibility that the entire system can enter another (undesirable) regime.[Holling 2002] Both aspects of resilience are critical to the performance of an ALS which typically has dual goals of shortening the time to return to nominal behavior and to avoiding a state of critical failure.

The closed and semi-closed natures of regenerative systems pose unique problems in measuring robustness or resilience. Typical approaches such as equating resilience with the quantity of unused resources that could be used to recover from a fault are unsuitable in closed systems because increasing the resources held in a particular buffer necessarily reduces the amount held elsewhere. This makes it hard to identify the states or operating modes that have the greatest capacity to recover from failures. Furthermore, resilience is a property not just of the hardware that makes up a system, but also of the way that hardware is controlled; two different control approaches can produce vast differences in the resilience of the same physical system. For these and other reasons outlined below, the analysis of the resilience or robustness of a system is best done dynamically, by measuring the behavior of the system over time as faults occur.

Resilience is a dynamic, rather than a static, property of the system. To see that this is the case, consider the definition of resilience as *"the persistence of specified system features in the face of a specified assembly of insults."* That is not to say that in a resilient system faults do not occur, but that when (some specified set of) faults or unexpected events occur, a resilient system can recover from them and return to normal operation. Thus the property of interest is the behavior of the system over time. Resilience involves determining whether normal system operation resumes after a fault, and whether this occurs within an acceptable amount of time (e.g. before the crew become ill). Dynamic analysis of the system is the only way these issues can be determined, and static measures such as equivalent system mass (ESM) cannot hope to fully capture these details. For some simple systems, exact dynamic analyses can be made directly by examining the system model. Because of the complexity and non-linear dynamics typical of advanced life support systems (ALSS), such analysis is most easily performed through simulation.

The next section examines the current ALS metric used to evaluate life support systems and shows new measures of resilience relate to existing techniques. The following section describes the specifics of our approach. The next describes the simulation test bed model of a generic WRS. Finally preliminary results and analysis of simulation data demonstrate how our proposed measures of resilience work in practice.

EXISTING METRICS AND RESILIENCE

We now briefly review the tool for computing the current advanced life support (ALS) metric, ESM. The ESM of a technology or subsystem is traditionally computed from static analyses, assuming nominal operation. ESM computations consider the mass, volume, power, cooling and crewtime requirements of the system under nominal behavior. The life support needs for volume, power, cooling, and crewtime are converted to units of mass to represent the required launch mass of the entire life support system (LSS). Mass units are used because the launch mass of a system is commonly correlated to mission cost. The reader is referred to Levri et al (2000) for a more detailed explanation of the static method of ESM computation.

As currently computed, ESM takes account of system robustness only in the time the crew spends on nominal maintenance, and the mass of spare parts needed for that nominal maintenance. The ability of a system to respond to off-nominal events is not captured in ESM. While ESM is a useful tool for evaluating the launch cost for a system that always performs nominally, as currently computed *it does not measure resilience to faults.* One of the main points in this paper is that *resilience is a dynamic property of a system as a whole*, and as such, is not adequately measured by metrics of nominal operation via static computation.

Valid ESM comparisons require that systems *"satisfy the same life support product quantity, product quality, reliability and safety requirements. In situations where product quality is somewhat subjective or the level of safety or reliability is not well defined, the researcher's expertise on those issues must be used to estimate appropriate requirements and relevant adjustments in ESM"* [Levri, et al, 2000]. Because ESM requires that two candidate systems must be equally resilient for a valid comparison, the development of explicit measures of resilience will complement and enhance this measurement.

Although modifications to the methods used to compute ESM could, be developed to consider dynamic, off-nominal operation, there are other roadblocks to capturing a system's resilience with ESM. *ESM is unable to reflect the improvement in a system that is made by simply changing the controls approach, without changing the mass, volume, power, cooling or crewtime needs.* In other words, if system A and system B are

equivalent in mass, volume power, cooling and crewtime needs, but system A is, in general, more resilient to off-nominal events, this advantage is not reflected in the ESM measure. Thus, progress made in the ALS Project in using advanced control approaches are not necessarily reflected in the ALS metric. Attempts should be made to apply the same rigor used in estimating ESM for two competing technologies to analyzing their robustness, resilience, and ability to meet critical performance requirements under off-nominal operating conditions.

A reliability measure that is often used is the *mean time before failure* (MTBF). MTBF is based on failure probabilities for each component of a system aggregated across for all system-critical components. Redundant components correspondingly reduce the MTBF for the whole system [Jones, 1999]. Unfortunately, for a complex system this estimate may be wildly inaccurate since it assumes a very simplistic relationship between component failures and overall system failure. In particular, it ignores interactions between non-critical system failures. In the simple WRS modeled in this paper, a series of bed breakthroughs, which are part of nominal system operation, can lead to increases in contaminant concentration in greywater tanks, which in turn cause more breakthroughs and potentially critical system failures. The MTBF of the beds does not, in isolation from additional data about the state of the system and the control system in place, provide a useful estimate of the probability of system failure.

Estimating the MTBF for the system as a whole in the presence of multiple, randomly injected, component-level faults is, in fact, one aspect of measuring a system's resilience. Ultimately, the complex interaction of faults, controls, and system dynamics, rather than simple fault probabilities, determines resilience.

METHODS

COMPUTATIONAL MODELING

Our approach to measuring resilience utilizes data from repeated simulation of a computational model of the system. Because resilience is an inherently dynamic property of a system, static measures that don't consider off-nominal behavior or account for control system responses to faults do not adequately capture resilience. Running extensive simulations of a model allows us to observe the dynamic effects of and the interactions between component faults, control system decisions and random variations in system inputs. This data is then analyzed to test the predictive power of different measures of system resilience.

To measure the resilience of a system by this method we develop the following:

1. A system model that exhibits both nominal and off-nominal system behavior and includes the uncertainty inherent in the system, for example in the amount of resources used by the crew of a life-support system or the contaminant capacity of a filtration bed;
2. A control system that handles both nominal and off-nominal operating conditions;
3. A fault model that describes the likelihood of each possible system fault occurring (possibly as a function of the system state);
4. An evaluation method for comparing the simulated performance of the system to the corresponding life support goals.

A single simulation of the model starting from a given initial condition is called, prosaically enough, a run, and the sequence of states that the system passes through during the entire run is its trajectory. While technically the state of a system is defined to be the minimum set of information required to derive the subsequent system behavior, here the term 'state' and 'state data' refer all of the descriptive data gathered during a run, while 'system performance' and 'performance data' refer to the much smaller set of simulation data that directly reflect the ability of the system to achieve critical life support goals. Examples of state and performance data for this work are shown in Tables 1 and Table 2 which appear below in the section describing statistics and data collection.

A model for examining system resilience needs to be detailed enough to have non-trivial dynamics and faults that occur at the component level, yet simple enough that its nominal behavior is well understood and that intuitive notions of system resilience can be confirmed with data. For example, it seems reasonable to assume that a water recovery system where the filtration beds have a higher capacity to remove contaminant is, with all other design elements and control decisions being equal, more resilient than one with a lower bed capacity. The system performance statistics from simulations confirm this rule-of-thumb notion of resilience, allowing us to use simulation data for systems that differ only in bed capacity to propose and test measures of system resilience. The measures of resilience are then compared to data for the same systems with induced probabilistic faults, to see if the measures are useful predictors of system performance in the presence of unanticipated faults.

A major strength of this approach is that it provides data about off-nominal behavior and examples of situations in which the system failed. Analysis of these can give valuable insights into the weaknesses of the system. Dynamic simulation-based data can be used to find the total probability of failure given the fault model or to determine the most probable trajectory of the system that leads it to fail. This analysis helps system designers identify the most vulnerable parts of the systems, not in terms of the probability of the underlying fault but in

terms of its dynamic effect on system performance. In addition, the data can support qualitative statements about the performance of the system, such as "the system is resilient to all single faults that were simulated." Finally, this approach allows direct comparison of the effects of two different control strategies applied to the same hardware or the effects of a particular hardware change on the performance of the system.

PROPOSED MEASURES OF RESILIENCE

We begin by describing two relatively simple approaches to measuring resilience, *summary statistics*, that describe performance over a set of runs with a single number, and *correlated system characteristics*, where resilience is measured indirectly by examining other features of the system that are intuitively connected with resilience.

Simple summary statistics such as the proportion of the runs in which the system continues to perform normally provide a starting point for examining resilience and a point of comparison for more sophisticated measures. The resilience of two different systems to a given set of faults can be simply compared using statistics that summarize how often each system fails on a set of runs including those faults.

Another attractive approach is to hypothesize that there are certain characteristics of the system, perhaps identified by a domain expert, that have some intuitive correlation with overall system robustness, but are easier to measure. These characteristics might provide useful measures of resilience. In the WRS model presented below, we might expect that the amount of reserve water in the tanks or the number of times contaminated water reaches the potable water tanks averaged over a number of runs, would be correlated with resilience. A system with lower amounts of reserve clean water is intuitively less resilient than one with greater reserves, and we might expect this relation show up in the summary statistics as a higher number of successful runs when there is more reserve clean water.

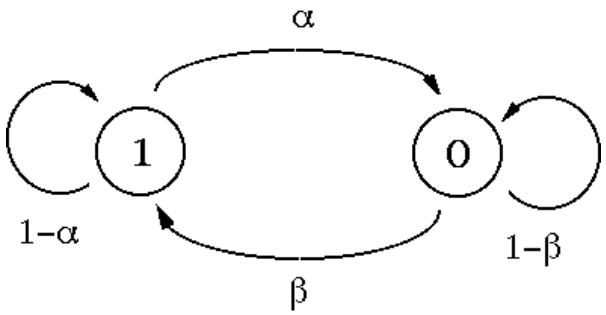
Summary statistics and correlated system characteristics can be used to compare the performance of competing systems, but may not provide a lot of information about design changes that could be made to improve a system's resilience. In addition, correlated system characteristics are only useful for comparing similar systems and control strategies. For example, a control strategy specifically tuned to work with lower overall water availability may be more resilient than other strategies even though it typically has lower water reserves.

A more sophisticated approach, which we are currently investigating, would assign a probability to each run that weights the data by the likelihood of the underlying faults and system inputs that produced it. Runs with very unlikely faults, or multiple faults, would have low probabilities associated with them, while runs with no faults, or single common faults would have higher probability. Now rather than using the proportion of successful runs to evaluate resilience, we can talk about the total probability mass of the successful runs, which should provide a more accurate estimate which of two systems is better. It may also help identify which design modifications are most likely to improve system resilience, since the highest probability runs with unsatisfactory performance are the most likely events that the system can't recover from.

MEASURING RESILIENCE WITH MARKOV CHAINS

In this paper we propose a novel approach to measuring resilience based on the parameters of a Markov chain model of system performance. A Markov chain is a concise way of describing the probabilities that the system will move from state to state[2]. Figure 1 shows a two-state Markov chain model of a system where state 1 represents normal system behavior and state 0 represents abnormal behavior. In the simulation model developed here, state 1 represents 'astronaut demand for clean water is being met' and state 0 represents 'astronaut demand for clean water is not being met.' In other life support applications, the states might represent acceptable versus higher than acceptable levels of CO_2 in the atmosphere.

Figure 1: A two-state Markov chain of system performance



Suppose that the system is currently in state 1. There are two possible outcomes in the next time step: the system can stay in state 1 or transition to state 0. The parameter α is the probability that the system moves to state 0, given that it is currently in state 1. Conversely, $(1 - \alpha)$ is the probability that the system stays in state 1, given that it is currently in state 1. β is the probability that the system moves to state 1, given that it is currently in state 0 and $(1 - \beta)$ is the probability that the system stays in state 0, given that it is currently in state 0.

Estimating the parameters α and β from the data collected from individual runs or aggregated over multiple runs is straightforward: simply count the number

[2] The key feature of Markov chain models is that the future state depends only on the current state, not on the entire history of the system up to that point. However, the validity of this assumption is not critical to its usefulness as a measure of robustness—what matters is that the Markov chain model provides a reasonable facsimile of the system behavior being analyzed.

of times that each event occurred [(1 -> 1), (1 -> 0); (0 -> 0), (0 -> 1)] and divide by number of times the system was in the relevant state. In other words, from an empirical point of view α is simply the number of times that the system changed from state 1 to state 0 divided by the total number of times the system was in state 1. Note that the each run must be broken up into discrete time periods to gather this data, and that the value of α and β are not independent of the choice of period.

A high β indicates that even if the system enters the abnormal state, it is highly likely to return to the normal state. because the system is more resilient than one where beta is low. Similarly, a low α is desirable because it indicates that the system is unlikely to enter the abnormal state in the first place. Resilient systems, that is, systems with the ability to successfully recover from faults, are likely to have high values of β and low values of α. Consequently, our proposed measure of resilience is simply β/α, which eliminates the units of time from the measure. This is a summary statistic or aggregate measure of system resilience over a given run or a specified length of time, consequently, the individual runs must be long enough to provide adequate data for estimating the parameters for the measure to be provide meaningful information.

A HYBRID COMPUTATIONAL MODEL FOR MEASURING RESILIENCE

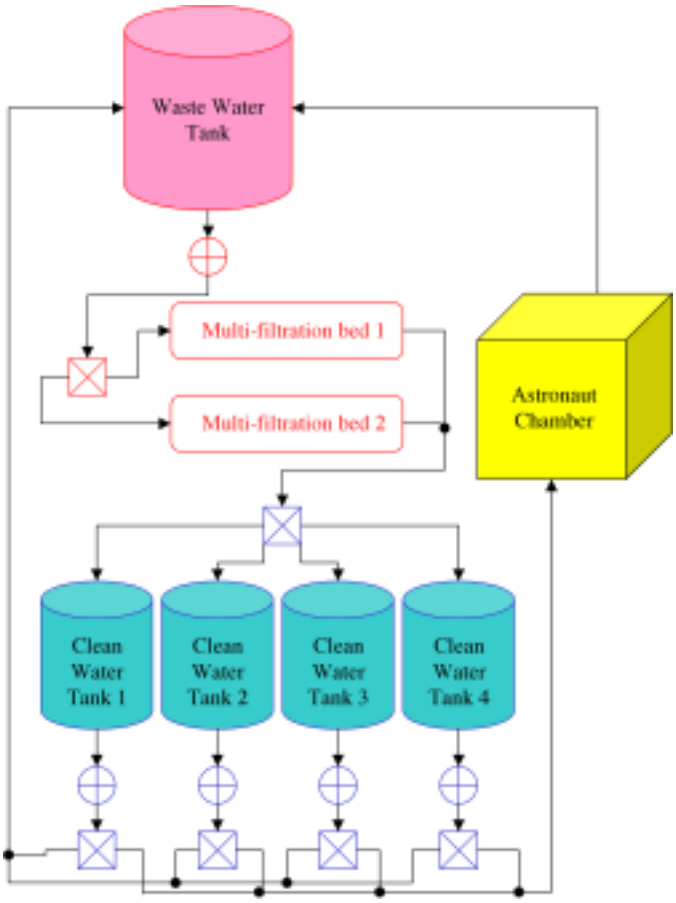
A model of a simplified WRS was developed to measure and analyze the resilience of regenerative life support technologies. The goal in designing the test system was not verisimilitude to a particular WRS, but rather, the creation of a relatively generic example of a regenerative life support subsystem constrained by conservation of mass. This simplified model provides a computational test-bed for developing methods to identify and measure resilience early in the design stage.

A GENERIC WATER REVITALIZATION SUBSYSTEM

The system consists of four main elements as shown in Figure 2:

1. *Astronauts*, who demand clean water for hygiene and drinking and introduce contaminant into the system. Detailed information on water use appears in Table 7 in the appendix.
2. *Waste Water Tank* (WWT), which receives greywater from the astronaut chamber and which control the variable flow of water to the filtration beds. The WWT also receives greywater directly from the potable water tanks when sensors indicate unacceptable water quality.
3. *Filtration Beds,* which remove contaminant from the greywater. Two beds operate in parallel with water flowing over one bed at a time. Beds remove 100% of inflow contamination until their capacity is reached. When the contaminant in the beds exceeds capacity they 'break through' and remove no contaminant (a step function). After breakthrough occurs, beds are regenerated to a new, randomly chosen capacity. The regeneration process takes one hour.
4. *Potable Water Tanks* (PWT), which receive clean water from the filtration beds and hold it during testing. There are four parallel tanks. Only one tank can be filling at a time, but more than one tank can drain at a time, either to the astronauts or directly to the WWT. Note that the outflow valve from the PWT controls the flow of clean water into the astronaut chamber, though its setting is determined by astronaut demand. Tank sizes and other parameters are detailed in Table 8 in the appendix.

Figure 2: Schematic of Generic WRS Model



There are two parallel mass flows: water and a generic contaminant. The schematic in Figure 2 illustrates these mass flows. Energy flow in the system (which would be added by pumping and lost by location change and friction in a physical system) is not explicitly modeled. In other words, mass flow is modeled but energy flow is not. Flows from element to element are controlled by switch valves (squares with x's), which select the outlet or destination of the flow, and variable flow valves (circles with crosses), which control the rate of flow out of a tank.

A separate control and sensor system collects data and controls the valve settings, the choice of filtration beds,

and the potable water tank assignments. The sensors record the quality of the water in each of the PWTs. The key control decisions are: 1) the setting of the outflow valve on the WWT; 2) the choice of bed in use; 3) the potable water tank being filled and being drained. In the baseline control system, the variable setting for the outflow valve on the WWT depends primarily on the WWT tank level. The change of filtration bed in use is based on a fixed schedule, and the PWT fill and drain completely before switching occurs. Additional system specification data appears in Tables 7, 8 & 9 in the appendix.

The quantity of water demanded, the capacity of the beds after each regeneration cycle, and the time that faults occur are generated randomly at the start of each simulation, but the simulation itself is deterministic. The flow of water and contaminant through the system is continuous, but many control decisions and fault occurrences are discrete, hence the WRS is a hybrid system. The simulation is implemented in Simulink, a hybrid system modeling language that works in conjunction with Matlab.

FAULTS IN THE TESTBED WRS

The simulation model is a tool for measuring resilience to component faults. Faults, along with possible control system responses and associated performance failures, are detailed in Table 9 in the appendix. For clarity of exposition, "faults" refer to the unanticipated failure of individual components while "failure" refers to performance failures of the WRS such as not providing clean water to the astronauts or providing untested water. While many potential faults can, and in a real system should, be compensated for with the appropriate control system response, in this paper we allow faults to occur in order to observe the off-nominal system behavior and resulting performance failures. (In other words, none of the control system responses detailed in Table 9 are actually implemented in the simulations presented here; they are merely listed as examples.)

Multiple faults can also interact, reducing the resilience of the system even more dramatically than a single, critical fault. A sensor fault makes it more likely that dirty water is sent to the astronauts, and can be compensated for by decreasing the amount of time before bed regeneration. In fact, anything that increases the number of PWTs that get filled with contaminated water also increases the probability that contaminated water is sent to the astronauts, as does any valve failure that prevents the control system from correctly routing water through the system. In addition, these faults occurring simultaneously can interact to dramatically increase the probability that the astronauts receive contaminated water. The discussion below, for example, explains how an interruption of water service can make future bed breakthroughs more likely.

Note that we do not consider bed breakthroughs as 'injected faults', as they occasionally occur even in systems with high absorptive capacity which do not experience any performance failures The number of bed breakthroughs that occur depends on the absorptive capacity of the beds and the length of time before bed regeneration. A decline in the average bed capacity or the failure of a bed to regenerate are, however, considered faults. Tank overflows are considered performance failures, not faults[3].

System design is often an incremental process in which the robustness and other characteristics of the system is gradually improved over a series of iterations. The robustness measure used can also be expected to change over these iterations. For example, during initial design, a system might be optimized only against ESM or some other static measure. Once a satisfactory design has been reached, dynamic analysis of the system performance in nominal operating modes might be used to further refine the system. Finally, a robustness measure such as the ones we are proposing can be used to dynamically analyze system performance in off-nominal scenarios, possibly even with different resilience metrics being used as the evolution of the system continues. These steps should be iterated as necessary.

In a similar, incremental manner, in the phase of robustness measurement, the control system should also begin with no failure prevention strategies included. Measuring the resilience of this stripped-down system gives information about the most important faults, and system changes designed to prevent those failures can then be added, the new system's resilience can now be tested, and so on. This process ensures that the control system is not needlessly complex by only including control responses to faults that actually occur in practice and to which the system is not already robust.

SIMULATION DATA & ANALYSIS

First we demonstrate the nominal behavior of the system, that is, the system behavior without injected faults, through example simulations of three possible system performance scenarios. Next we present summary statistics for system performance under different specifications of bed capacity. We use this data to propose different measures of system resilience. Finally data for the same systems with probabilistic faults injected is used to verify that the measures of resilience are in fact useful predictors of system performance in the presence of unanticipated faults.

[3] Note that all tanks have overflow valves designed to prevent them from holding more water than their capacity, excess water flows onto the 'floor' and is not returned to the system in this simple example. We do not consider overflow valve failures which would cause water to back up in the system.

NOMINAL SYSTEM BEHAVIOR: EXAMPLES

- Case 1: some bed breakthroughs, no interruptions of water service;
- Case 2: some bed breakthroughs, some non-life threatening interruptions of water service;
- Case 3: bed breakthroughs leading to complete system breakdown and loss of crew.

In all three cases the filtration beds have a 'medium capacity' (see Appendix for specification details).

Case 1: No Performance Failures

We start with Case 1, where some filtration beds break through but clean water is supplied continuously to the astronauts. In Figure 3, the vertical lines (blue) show the filtration bed breakthroughs. The signal 1 indicates a breakthrough and 0 indicates no breakthrough, consequently, when a breakthrough occurs it appears as a vertical line. There are 8 breakthroughs in this run. The red line indicates water availability with 1 indicating that the demand for clean water is met and 0 indicating that no clean water is available. Figure 4 shows the clean water reserves in the system. Figure 5 shows the dynamics of the four potable water tank levels, with each tank level indicated by a different line (red, blue, light blue, green). When tanks are draining and filling rapidly the lines indicating tank levels are more vertical and closer together. Less dense portions of the graph with less steeply sloped tank levels indicate times of lower water usage and slower flow over the filtration beds.

Figure 3: Case 1, Bed Breakthroughs and Water Availability

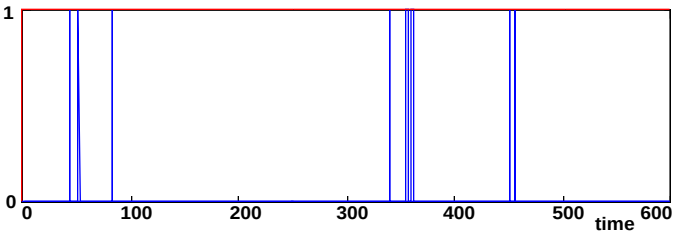


Figure 4: Case 1, Clean Water Reserves

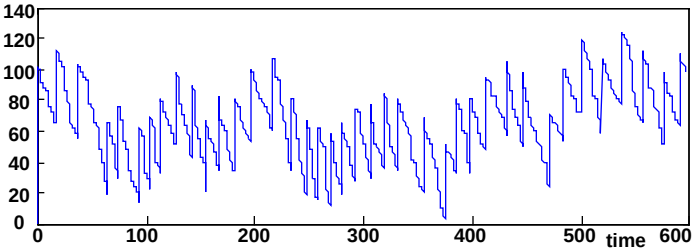
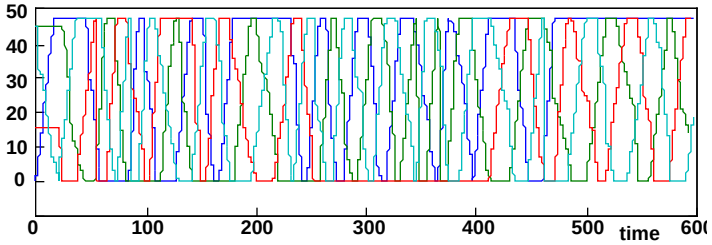


Figure 5: Case 1, Potable Water Tank Levels



Relatively low demand in the period (175–225), and the corresponding slow drainage of clean water from the tanks in that period (indicated by the sparse section of Figure 5), leads to an increase in clean water reserves. High demand causes low water reserves in the period (250-350), the subsequent breakthroughs that occur in the (325–375) range almost lead to an interruption in water service.

Case 2: Non-Critical Interruptions of Water Service

In Case 2 the system also has some filtration bed breakthroughs, but also experiences some interruption of water service. Figures 6, 7, & 8 present the data for this case. In Figure 6, the first vertical line (red) marked with an arrow indicates a short service interruption of 3 minutes, the next two vertical lines (red) marked with an arrow demarcate a more serious service interruption lasting 6 hours and 37 minutes. The unmarked vertical lines (blue) indicate bed breakthroughs. Note that the water shortage is caused by the need to empty multiple dirty PWTs in the period (275-300) caused by the earlier bed breakthroughs. The relatively dense colors in this period in Figure 8 result from this filling and rapid draining of tanks. There are 7 breakthroughs in this run. Note that fewer breakthroughs occurred in Case 2 than in Case 1, but that in Case 2 the breakthroughs resulted in service interruptions.

Figure 6: Case 2, Bed Breakthroughs and Water Availability

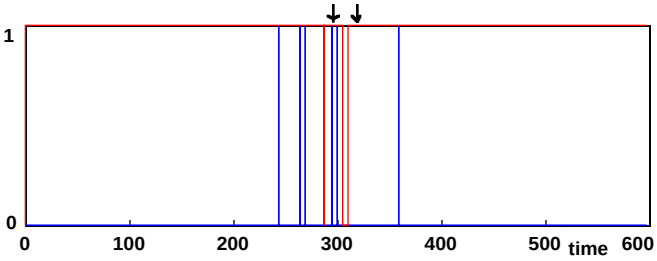


Figure 7: Case 2, Clean Water Reserves

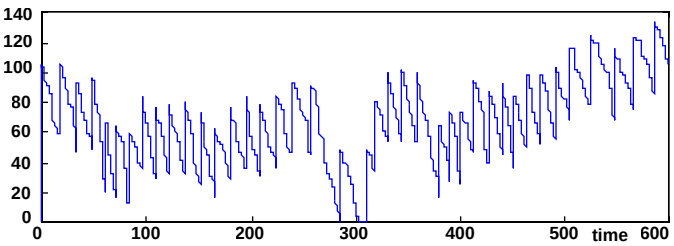
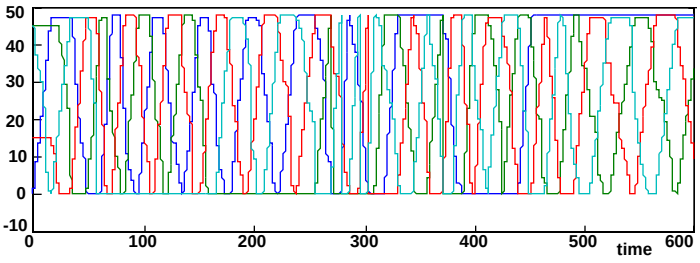


Figure 8: Case 2, Potable Water Tank Levels

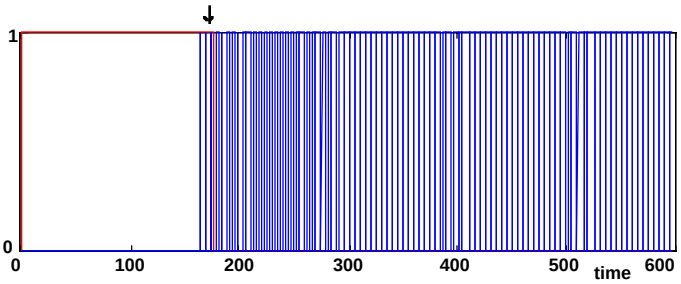


Case 3: Critical Performance Failure

The final case demonstrates how this simple WRS with a control system that does not compensate for component failures can fail completely. Figures 9, 10, 11 & 12 show the system breakdown around time 175. The vertical line (red) marked with an arrow shows the end of water service to the astronaut chamber. The almost solid swath of color in Figure 12 shows the frantic but ultimately unsuccessful filling and draining of the PWTs with contaminated water.

How does this critical failure occur? The key point is that the astronauts continue to produce contaminant even when no water is available[4]. Otherwise, a system with service interruptions would have less contaminant to process overall. This results in a higher concentration of contaminant in the WWT, shown in Figure 12 In the absence of a compensating control action like decreasing the amount of time between bed regenerations, this rise in the contaminant level makes additional bed breakthroughs more likely in the future. This failure propagation mechanism or positive feedback loop can, under certain conditions, lead to total system failure.

Figure 9: Case 3, Bed Breakthroughs and Water Availability



[4] In other words, the astronauts continue to sweat , and so on, even when no water is available for showering. The most realistic scenario would be for the contaminant produced during interruptions of water service to be stored in the astronaut chamber and then flushed out at higher rate once service resumes. For ease of programming the contaminant simply continues to flow on schedule to the WWT without the corresponding clean water flow. The effect on the system will be the same in both cases: the concentration of contaminant in the WWT will increase whenever the demand for water is not met.

Figure 10: Case 3, Clean Water Reserves

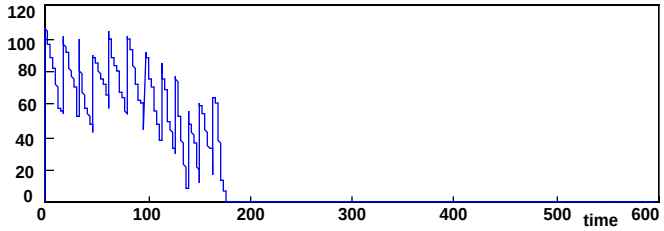


Figure 11: Case 3, Percent Contaminant in WWT

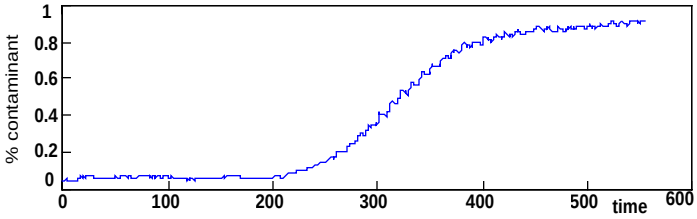
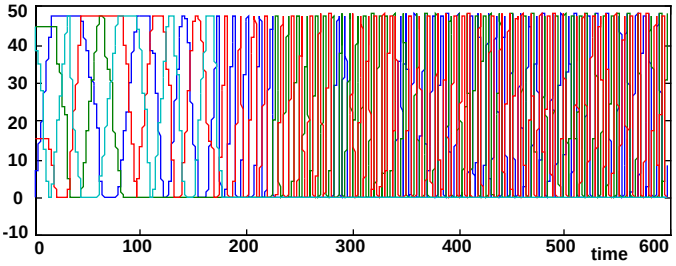


Figure 12: Case 3, Potable Water Tank Levels



SUMMARY STATISTICS AND DATA COLLECTED

The analysis and measurement of system resilience supports the overriding goal of ensuring crew safety and health. The WRS provides a reliable supply of contaminant-free water for astronaut hygiene and drinking. The possible performance failures that we consider, in rough order of severity, are:

1. interruption of water service for more than 72 hours
2. untested contaminated water sent to astronauts;
3. interruption of water service for less than 72 hours;
4. untested clean water sent to astronauts;
5. loss of system water through leaks or overflows.

Performance failures 2, 3 & 5 can only occur with injected faults, for example, faults involving sensors, valves or controls.

The key question we are trying to answer is: What data best measure or predict the ability of the system to avoid performance failures in the presence of faults? This suggest two subsidiary questions: What data should we collect and analyze in order to develop a measure of system resilience? What data are likely to be available in a functioning WRS?

Tables 1 and 2 summarize the data collected from the simulations. There are two basic types of data: time series data, which consist of instantaneous observations of the state of the system at each time point, and summary data, which consist of state and performance data aggregated over an entire simulation or a given length of time. (Of course, summary data are calculated from time series data.)

Table 1: State Data Collected	
times series data[5]	**summary data**
% contaminant in WWT	average % contaminant in WWT
WWT level	average WWT level
bed breakthrough status	total number of bed breakthroughs
% bed capacity used	average and/or total % bed capacity used
---	total number of bed changes
clean water reserves	average clean water reserves
contaminated water in PWT	total number of contaminated PWTs drained to WWT
untested water in PWT	---
astronaut demand for water	total astronaut demand for water
supply of water to astronauts	total supply of water to astronauts
---	total water processed
contaminant produced by astronauts	total contaminant produced by astronauts
---	total contaminant processed

Table 2: System Performance Data Collected	
times series data	**summary data**
% demand for water being satisfied	total % demand for water satisfied
---	number of water service interruptions
---	total length of water service interruptions
---	maximum length of water service interruptions
untested clean water flowing to astronauts	total quantity of untested clean water sent to astronauts
contaminated water flowing to astronauts	total quantity of contaminated water sent to astronauts
---	number of times astronauts received contaminated water
PWT, WWT tank overflowing or leaking	total water lost due to overflows or leaks

[5] Additional data collected includes the tank level, percent contamination, filling status, draining status and test status of each of the PWTs; the valve settings, the quantity of water in the astronauts; and the total water in the system.

The distinction between the two data types is useful for more than ease of presentation---it also corresponds to two distinct, complementary ways of thinking about resilience. Aggregate statistics collected over a long period time can be used to measure the overall resilience of a particular system. An aggregate or summary measure of resilience answers the question, on average, how well can this system recover from faults? Summary measures are particularly useful when comparing alternative system specifications.

Time series data, on the other hand, lead to an instantaneous or conditional measure of resilience that reflects not only the system's overall design but also its current state. An instantaneous measure of resilience answers the question, would this system be able to recover if a fault occurred right now? An instantaneous measure of resilience reflects the system's ability to recover from faults, conditioned upon the current state of the system. Conditional measures of resilience are useful in identifying the best nominal operating point of a system and may provide an early warning of system vulnerability before faults have occurred.

The Markov chain measure of resilience proposed above and most of the statistics presented here are aggregate measures of resilience and system performance. Future research will examine instantaneous measures of resilience.

NOMINAL SYSTEM BEHAVIOR:

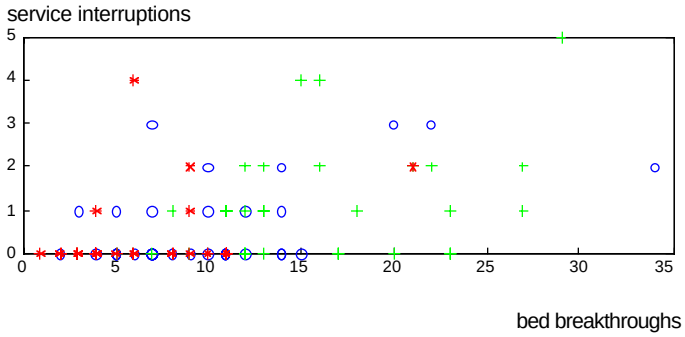
Table 3 presents performance data aggregated across runs, of systems with three different average bed capacities but no injected faults. The performance measures correspond to Cases 1, 2, & 3 from the previous subsection.

Table 3: Summary Data - Nominal Operation			
summary data	Average Bed Capacity		
	high	med	low
total number of runs of 1200 hours	27	34	45
percentage of runs with no service interruption	81 %	59 %	31 %
percentage of runs with some service interruption	19 %	35 %	38 %
percentage of runs with critical failures[6]	0 %	1 %	31 %

[6] We are aware that most self-respecting life support engineers would likely find the number of critical failures reported here unacceptable. More sophisticated control strategies, for example, switching beds immediately after the sensors indicate that a PWT has been contaminated, would perhaps lead to better system performance and recovery from certain kinds of faults. However, the goal is not to optimize the current system but rather to develop methods for measuring and analyzing the resilience of the system in the presence of probabilistic faults, which requires examination of off-nominal behavior.

Figure 13 plots the number of bed breakthroughs versus the number of service interruptions for each of the 93 runs without critical failures. Each asterisk (red) represents one 1200 hour run with high bed capacity, circles (blue) are runs with medium capacity and plus signs (green) are runs with low capacity. The weak correlation between bed breakthroughs and service interruptions shows that bed breakthroughs are not, on their own, good aggregate predictors of system performance despite the fact that they are the only source of performance failures in these simulations. This emphasizes the importance of dynamic analysis in understanding resilience.

Figure 13:
Number of Bed Breakthroughs versus Number of Service Interruptions.
+ (green) = low bed capacity, o (blue) = medium, * (red) = high)



Recall the parameters of the Markov chain described above: α represents the probability that system goes from the desired state of providing clean water to the undesirable state of not providing water and β represents the probability of moving out of the state of no water service to the desired state of providing clean water. Despite the rich array of data collected the parameters (α, β) and the related measure of resilience, β/α, described above can be derived from a few simple statistics: the number of times that the system provided and failed to provide water, the total length of time the system failed to provide water, and the total length of time under consideration[7]. Note that the measure is not defined on individual runs or periods of time where no performance failures occurred. A two state Markov chain is also not designed to model the behavior of a system with critical failures[8].

Figure 14, found in the Appendix, shows the α and β parameter estimates for individual runs with at least one service interruption and without critical failures. The inclusion of critical failures would dramatically change the parameter estimates for the medium and low bed capacity. Measures derived from data without critical failures, taken by itself, will tend to overstate the resilience of systems with critical failures. However, including the data for performance failure long after the system has ceased functioning would similarly skew the results[8].

The ideal point of system performance is found in the lower right hand corner of the graph, where a low α indicates that the system is unlikely to lose water service and a high β indicates that the system is likely to regain service quickly after an interruption. Moving diagonally from the upper left hand corner to the lower right hand corner represents an unambiguous improvement in system performance.

In the upper right hand corner of Figure 14 the system is relatively likely to experience an interruption in water service but is also likely to regain service quickly. On the other hand, in the lower left hand corner of Figure 14 the system is unlikely to experience a service interruption but also unlikely to recover from an interruption once it occurs. Evaluating system performance along this diagonal axis is more ambiguous. Are lots of short service interruptions considered to be less serious than fewer longer interruptions? If so then a high β should be given more weight than a low α. On the other hand, if continuity of service is critical then a low α should be the primary consideration.

We combine these two competing aspects of system performance into a single measure, the ratio of the two parameters: β/α. Figure 15 presents a contour plot indicating how the proposed measure rates different combinations of α and β. An alternative function of the two parameters which weights the tradeoff between the two parameters differently could be developed to reflect different performance goals.

Table 4 presents the Markov chain parameters estimated for the aggregated data for all runs taken together, without and with the critical failures included. Note that α, the probability of losing water service, is typically much lower than β, the probability of regaining water service after it is lost. This indicates that the system is typically in the desirable state of providing water service. When α is greater than β then the system spends more time in the undesirable state of not providing water.

Table 4:
Markov Parameters - Nominal Operation

Mea-sure	Bed Capacity				
	high	med	low	med*	low*
α	0.0003	0.0005	0.0009	0.0006	0.0013
β	0.2035	0.2984	0.1792	0.0156	0.0058
β/α	658.22	544.65	202.05	24.52	4.49
*includes data from runs with critical failures					

What does the measure β/α say about the resilience of

[7] For these calculations time is measured in hours. As mentioned previously, the estimates of α and β are not independent of units of time. The length of time over which the parameters are estimated is either a single run or a group of runs aggregated together. It could also be a subset of a particular run.

[8] An empirical compromise might be to eliminate the data from the time that system is determined to have failed, after 72 hours, to the end of that run.

the three systems with different bed capacity? The proposed measure of resilience confirms the intuition that a system with higher bed capacity should be more resilient than one with a lower bed capacity. However, the difference between the measured resilience of the low and medium capacity is large, but the difference between that of the medium and high capacity is quite small (excluding the runs with critical failures). This reflects the similar performance of the two systems, for example, there is no water service for 0. 1519% of the time with high capacity versus no water service for 0.1836% of the time with the medium. This would indicate to a system designer that the marginal benefit of improving the average bed capacity the first 0.05 increment from 1.00 to 1.05 is large, whereas an additional 0.05 increase provides only a small increase in the reliability of the system[9].

SYSTEM BEHAVIOR WITH A VALVE FAULT

Next we consider the behavior of the system and the corresponding measures of resilience when a valve fault occurs. Again, there are three different bed capacities. The fault occurs in the switch outflow valve of PWT 4, which sticks in the position that routes water to the astronaut chamber. Consequently, if the tank is filled with greywater as the result of a bed breakthrough there is no way to drain it to the WWT and it cannot be used. The PWTs are now three tanks in parallel.

Table 5 presents the summary data for this scenario. As expected, the percentage of runs with no interruptions in water service declines in all cases. The high bed capacity system experiences one critical failure in this data set, as does the medium capacity system. One anomalous but robust result is that the number of critical failures in the system with low bed capacity declines.

Table 5: Summary Data - Valve Fault

summary data	Bed Capacity		
	high	med	low
total number of runs of 1200 hours each	25	33	36
percentage of runs with no service interruption	60 %	36 %	19 %
percentage of runs with some service interruption	36 %	61 %	64%
percentage of runs with critical failures	4 %	3%	17 %

Table 6 presents the estimates of the Markov chain parameters and measures of resilience for data aggregated across all runs, both without and with the runs that experienced critical failures. In all cases the systems with the fault have lower measured resilience than the corresponding systems without faults, and the relative reductions in resilience are the same for the different system specifications. (The estimates based on data that included critical failures deviates slightly from the previous pattern --- the estimates of β for the high and medium systems are heavily influenced by the amount of time that the system fails to provide water after the one critical failure in each sample occurs leading to a reversal of their measured resilience.)

Table 6: Markov Parameters with Valve Fault

measure	Bed Capacity					
	high	med	low	high*	med*	low*
α	0.0012	0.0018	0.0040	0.0012	0.0018	0.0013
β	0.2278	0.1688	0.1643	0.0299	0.1183	0.0058
β/α	191.99	95.77	41.54	24.59	66.90	4.50
*includes data from runs with critical failures						

The usefulness of the proposed measure can and should be explored further using additional data from a wider variety of faults. The predictive capacity of the measure can them be compared to the simulated system performance in different scenarios.

CONCLUSION

This paper argues that crew safety depends on the resilience of the ALS system; its ability to achieve crucial performance goals in the presence of unanticipated faults. Static measures of system quality are unlikely to capture this important characteristic. We propose a method for analyzing system resilience early in the design process, rather than waiting until detailed design information is available. We develop a computational model of a WRS that serves as a testbed for different measures of resilience, and propose a novel measure that uses the parameters of a Markov chain representation of system performance to measure resilience. The measure performs well in this initial analysis of off-nominal system behavior.

ACKNOWLEDGMENTS

The authors thank William Sethares, Adam Sweet and Udo Toussaint for helpful comments. This work was supported by the Director's Discretionary Fund of the NASA Ames Research Center.

REFERENCES

1. Allen, C. R. Ecosystems and immune systems: hierarchical response provides resilience against invasions. Conservation Ecology 5(1): 15. 2001. URL: http://www.consecol.org/vol5/iss1/art15
2. S. Carpenter, B. Walker, J. M. Anderies and N. Abel. From Metaphor to Measurement: Resilience of What to What? Ecosystems 4, 765-781, 2001.

[9] Keeping in mind that a more sophisticated control would likely avoid the critical failures.

3. A. Drysdale. Metrics and System Analysis. SAE Technical Paper 981746, 28th International Conference on Environmental Systems, 1998.
4. B. Holling. Definitions of Resilience. 2002. http://www.sustainablefutures.net/resilience/resilienceDef.html
5. H. Jones. Multiple Metrics for Advanced Life Support. SAE Technical Paper 1999-01-2079, 29th International Conference on Environmental Systems, 1999.
6. J. Levri, D. Vaccari and A. Drysdale. Theory and Application of the Equivalent System Mass Metric. SAE Technical Paper 2000-01-2395, 30th International Conference on Environmental Systems, 2000.
7. Santa Fe Institute. document ref. # RS-2001-009 http://discuss.santafe.edu/robustness/

CONTACT

Julie Levri, MS 239-8, NASA Ames Research Center, Moffett Field, CA 94035-1000 jlevri@mail.arc.nasa.gov

ACRONYMS & ABBREVIATIONS

ALS: Advanced Life Support
ALSS: Advanced Life Support System
ESM: Equivalent System Mass
LCC: Life Cycle Cost
MTBF: Mean Time Before Failure
PWT: Potable Water Tank
WRS: Water Revitalization System
WWT: Waste Water Tank

APPENDIX

Table 7: Astronaut demand for water use and contaminant produced[10]

event	# per day	mean length	range	flow rate	% cont	total use	cont
shower	6	6 min	3 – 9 min	48.0 lit/hr	0.5	28.80	0.1440
hand/face	6	3 min	1.5 – 4.5 min	54.4 lit/hr	2.0	19.32	0.3864
teeth	12	2 min	1 – 3 min	10.8 lit/hr	7.0	4.32	0.3024
drinking	24	1 min	0.5 – 1.5 min	44.4 lit/hr	0.0	17.76	0.0000
urine	24	1 min	0.5 – 1.5 min	44.4 lit/hr	20.0	17.76	3.5440

Table 8: Tank, valve, sensor and filtration bed parameters

tanks	size	type	max flow	controlled by
WWT	200 lit	---	---	inflow determined by outflow of dirty water from astronauts and draining of dirty water from PWTs;
PWT	50 lit	---	---	inflow determined by outflow from WWT; tank filling chosen by control system, 'untested' tank with highest level filled first;
valves				
WWT outflow	---	variable	5 lit/hr	valve setting is proportional to tank level, nominal outflow = 3, higher if level > 50% max = 5 when level > 90%, min = 2;
filtration bed inflow	---	switch	---	beds switched every 5 hours
PWT outflow to astronauts	---	variable	based on demand	astronaut demand
PWT outflow to astronauts	---	switch	---	tank draining chosen by control system, 'clean' tank with lowest level drained first
PWT outflow to WWT	---	variable	50 lit/hr	'dirty' tanks drain to WWT at the max flow rate immediately after sensors detect contamination
sensors	Attached to each PWT, 'test status' consists of untested/being tested, clean, dirty/contaminated. An empty or filling tank is 'untested'. Tests take one hour.			
filtration beds	Capacity after regeneration is uniformly distributed over the ranges: [1,2] (low capacity); [1.05, 2.05] (medium capacity); [1.1, 2.1] (high capacity.)			

[10] The lengths of the water-using events are uniformly distributed over the given range. Astronaut data is based on 6 astronauts using on average a total of 70.2 liters of water per day and producing on average 4.3768 liters of a generic contaminant per day at a rate of .1824 liters per hour. Note that astronauts do not respire or perspire water.

Table 9: Component Faults, Control Responses and Performance Failures			
system element	faults	possible control system responses	associated performance failures
WWT	outflow valve stuck open	drain excess untested water from PWT back to WWT, if necessary, to prevent PWT overflow	decline in clean water reserves or service interruption if valve stuck at low level
	outflow valve stuck closed	limit supply of water to astronauts to drinking water, if necessary, to prevent WWT overflow	decline in clean water reserves or service interruption
	tank leaks	---	total water in system declines
filtration beds	complete failure of bed, cannot be regenerated	adjust WWT valve control so no water flows while remaining bed regenerates	decline in clean water reserves or service interruption
	switch valve stuck	same as above	same as above
	decline in average bed capacity after regeneration (not directly observable)	decrease amount of time before bed regeneration	decline in clean water reserves or service interruption, increased chance that dirty water in PWT tanks is sent to astronauts
	bed fails to regenerate during one cycle (not observable)	---	same as above, but much less severe
PWT	inflow switch valve stuck	interrupt water service while tank fills and tests, set WWT outflow valve to maximum setting, decrease amount of time before bed regeneration to compensate for increased percent contamination in WWT	service interruption, tank overflows if fault not observed, increased chance that dirty water in PWT tanks is sent to astronauts
	outflow switch valve stuck	stop filling tank, tank out of service	decline in clean water reserves or service interruption
	variable outflow valve stuck	stop filling tank, use switch valve to route water to WWT, tank out of service	decline in clean water reserves or service interruption, increased chance that dirty water in PWT tanks is sent to astronauts
	sensor failure false positives	regenerate beds more frequently	increased chance that dirty water in PWT tanks is sent to astronauts
	sensor failure false negatives	---	decline in clean water reserves or service interruption
astronauts	leave water running, increase water usage, decrease contaminant concentration	increase flow rate over beds	decline in clean water reserves or service interruption
	produce excess contaminant, same water usage, increase contaminant concentration	decrease time before bed regeneration	decline in clean water reserves or service interruption, increased chance that dirty water is sent to astronauts

Figure 14: Estimates of **β** plotted versus estimates of **α** for nominal operation without induced faults, data from runs with at least one service interruption and without critical failures.
+ (green) = low bed capacity, o (blue) = medium, * (red) = high

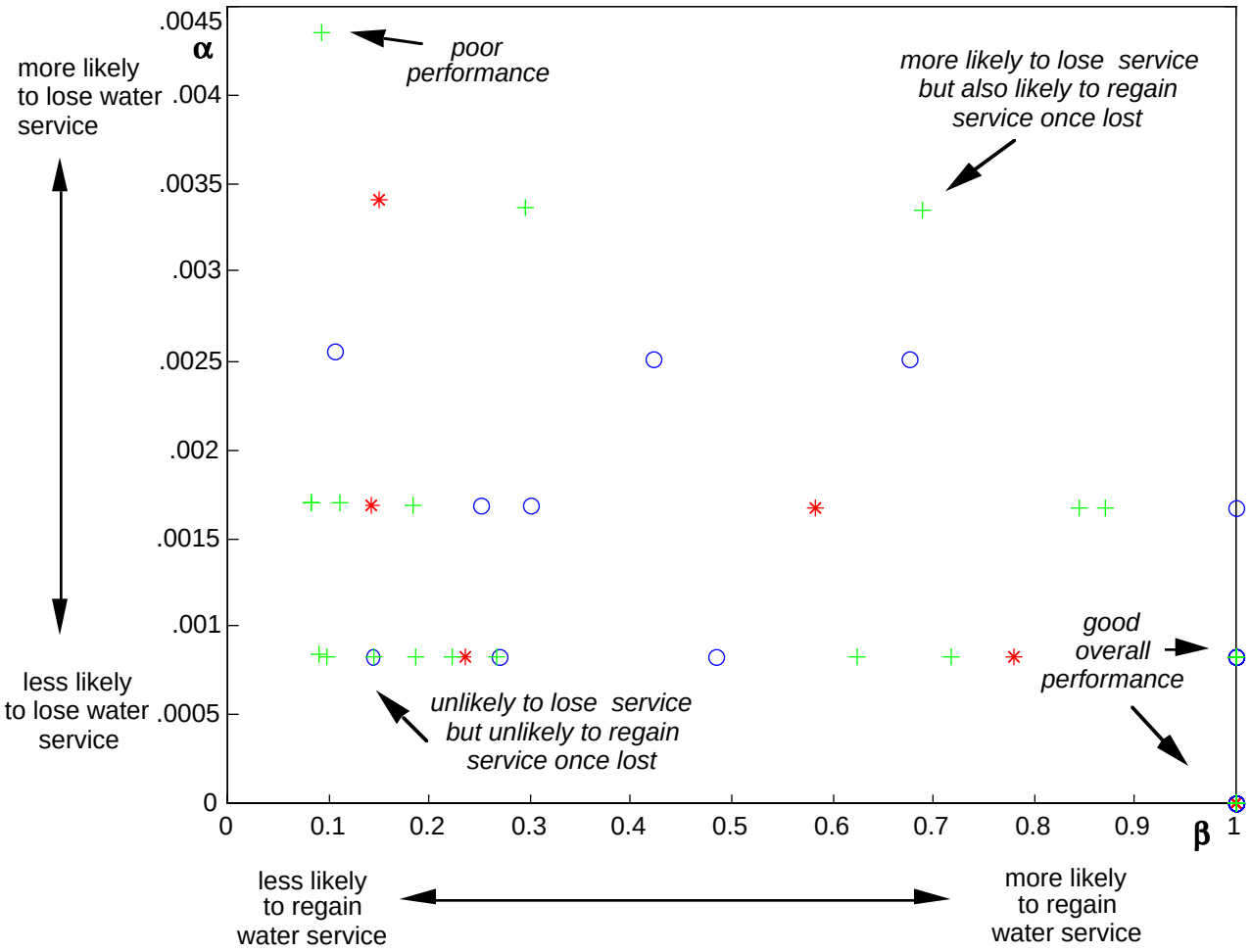


Figure 15: Contour plot of β/α with darkest areas representing poor performance and lightest areas representing good performance.