The System Modeling and Analysis of Resiliency in STEReO (SMARt-STEReO)

Sequoia Andrade*
HX5, LLC., Moffett Blvd, Mountain View, CA, 94035

Eleni Spirakis†
University Space Research Association, Moffett Blvd, Mountain View, CA, 94035

Daniel Hulse‡, Hannah S. Walsh§, and Misty D. Davies¶
NASA Ames Research Center, Moffett Blvd, Mountain View, CA 94035

NASA's Scalable Traffic Management for Emergency Response Operations (STEReO) project aims to leverage Unmanned Aerial Systems (UAS) and UAS Traffic Management (UTM) to improve asset coordination and overall emergency response. One application of STEReO is wildfire response, which is the focus of this research. In order to implement the operations described in the STEReO project, these additions must have tangible benefits and proven safety. To this end, the System Modeling and Analysis of Resiliency in STEReO (SMARt-STEReO) project constructs a simulation model, developed through the Python modeling and resiliency analysis package fmdtools. The model describes wildfire response operations, including current operational concepts and emerging concepts utilizing UAS as described in STEReO. While previous simulation models focus primarily on fire propagation with some models including emergency response intervention, SMART-STEReO evaluates the system performance and resilience benefits gained by the addition of UAS and UTM. Due to the novelty and complexity of the model, initial model verification and validation efforts are conducted and a detailed description of the model is provided. Preliminary results from experimental analysis on the SMARt-STEReO model indicate that when compared to current operations, the addition of UAS in wildfire operations results in improved response efforts, in terms of fewer acres burned, as well as improved system resilience in response to a given fault.

I. Introduction

Wildfire response has traditionally relied on relatively low-tech solutions, such as radiotelephony, to coordinate ground and aerial assets. In this setting, operators must listen to multiple radio frequencies and use this information along with line-of-sight visuals to coordinate their response. While these solutions are necessarily robust in remote environments, they limit strategic cross-organizational support. This includes limitations of the ability to deploy and effectively utilize aerial assets, such as Unmanned Aerial Systems (UAS). As aircraft become more advanced and new technologies such as drones and autonomous systems become available, there is potential to improve emergency wildfire response by using a new, more modern method of asset coordination that incorporates these technologies. It is expected that UAS will benefit wildfire response operations in a number of ways, including surveillance and logistics delivery. UAS traffic management (UTM) can improve coordination in these situations. However, because the application of these technologies in wildfire response operations is still relatively new [1], there is a need to understand and quantify the benefits and drawbacks of these systems.

Resilience is a growing concern for emerging aviation operations due to issues of known brittleness in autonomy and the safety criticality of such operations [2]. Autonomous systems, such as UAS, often perform best only under scenarios they are specifically designed to handle, resulting in "brittle" rather than resilient operations. Resilience, in contrast, is the inherent ability of a complex engineered system to mitigate hazardous scenarios from a comprehensive

*Engineer I, sequoia.r.andrade@nasa.gov
†Graduate Student, Stanford University, spirakis@stanford.edu
‡Computer Engineer, Intelligent Systems Division, daniel.e.hulse@nasa.gov
§Computer Engineer, Intelligent Systems Division, hannah.s.walsh@nasa.gov
¶Research Scientist, Intelligent Systems Division, Mail Stop 269-3, AIAA Associate Fellow, misty.d.davies@nasa.gov

(prevention and recovery) standpoint [3]. Because of the complexity and risk involved in wildfire response, there are many unforeseen contingencies that current emergency response operators must manage. These high-risk considerations include the unpredictable behavior of fire rate and direction of spread due to weather, terrain, and fuel, the need for timely coordination, and the human risk of both civilians and emergency response personnel. In these operations, faults and failures are inevitable and costly. Therefore it is important to design these autonomous systems in aerial wildfire response to prevent and adapt to faults [4, 5]. To enable safe and effective operations of systems involving autonomy, such as UAS and UTM technologies, it is therefore important to explicitly consider resilience in addition to performance to assess whether the technologies can adapt to contingencies as well as – if not better than – current operators. This research investigates how adding UAS and UTM into wildfire response impacts the performance and resilience of the response effort.

To this end, the "System Modeling and Analysis of Resiliency in STEReO" (SMARt-STEReO) project develops and utilizes an integrated model of wildfire propagation and response. The model enables the quantification of the additional system resilience gained from the use of UAS for surveillance and logistics delivery. While there are existing models of fire propagation [6–9], some of which also include the influence of suppression efforts [10, 11], there has been little research on the resiliency impact of emerging technology using modeling and simulation. The SMARt-STEReO model demonstrates how a wildfire response system contains fires by modeling aerial and ground assets, fault scenarios, and ground/weather conditions. This model uses the fmdtools Python package for simulation and resilience quantification [12], which enables the simulation of a systems response, and thus resilience, to fault scenarios under different response concepts (e.g., with and without different assets such as UAS). The model simulates the behavior and interactions of various firefighting personnel, including user-defined combinations of air tankers, helicopters, UAS, ground crews, and the resulting dynamic effect on fire propagation.

Due to the complexity of the model, it is first necessary to determine whether the scope of the model is appropriate for experimental analysis. As such, this study first examines model verification and validation prior to trends in the performance and resilience of STEReO technologies. The research questions (Q1-3) addressed are:

1. What is the appropriate scope of applications that the model is currently verified and valid for?
2. (a) Is there a difference in performance, in terms of metrics such as acres burned, when UAS is added into wildfire response?
   (b) If so, is this difference apparent in both nominal and off-nominal conditions?
3. (a) Does the addition of UAS in wildfire response increase resilience in the system in response to an individual fault?
   (b) If so, what is the effect size?

In addressing questions 2 and 3,the following hypotheses are developed and statistically tested (H2-H3):

2. (a) We predict the implementation of UAS will increase performance in wildfire response in both conditions.
   (b) The increase of performance will be at least equally, if not more, apparent in off nominal cases
3. (a) We predict that the addition of UAS will increase resilience in response to a given fault in the system.
   (b) We predict the effect size to be small to medium, however, the effect size will vary depending on the fault injected. The experimental effect size may not accurately represent the true effect given our modeling methodology and limitations to our model.

The following section provides background information on the current approaches to wildfire suppression, the proposed implementation of STEReO, current methodologies and approaches to resilience, and a description of fmdtools. Then, a detailed description of the model is provided. Next, to answer the first research question, preliminary model verification and validation efforts are presented. Finally, remaining research questions are experimentally explored through statistical analysis performed on results from Monte Carlo simulations on the model.

II. Background

A. Wildfire Response Operations

According to the U.S. Forest Service, there are an average of 7500 wildfires every year on National Forests and Grasslands, over half of which are caused by humans [13]. The top priority of wildfire response is safety of the public and of personnel [13]. In some cases, wildfires can be safely and effectively managed to reduce fuels and provide benefits to certain ecosystems [13]. However, left unchecked, wildfires can threaten human lives, private property, and natural and cultural resources [13]. To contain wildfires, strips of nonflammable material that inhibit the spread of the fire, known as fire lines or control lines, are constructed. Ground crews are responsible for the construction of fire lines, primarily using hand tools to dig to mineral soil where a line is desired. Many of the other wildfire response activities exist to support the construction of fire lines. Notably, aerial operations work to support fire line construction in several ways. Air tankers and helicopters perform water or retardant drops. Drops assist ground crews by providing them the opportunity to complete sections of line, to connect sections of line where line construction is difficult and slow, to cool off a section of line to allow ground forces to directly attack, to strengthen and reinforce control lines which may be too narrow to contain the fire, and to pretreat fuels in advance of line burning [14]. Additionally, aircraft can be used for crew and supply transport, reconnaissance, and aerial ignition [14].

At present, much of the technology wildfire response uses to communicate, coordinate, and commence operations has not significantly changed over the years. Typically, the only means of communication are with voice through radio and a visual line-of-sight. The response is managed by the incident commander, who identifies areas of interest and assigns ground or aerial crews to the targets. For aerial operations, an aerial supervisor receives the targets from the incident commander, decides the best methods to proceed utilizing the resources available, and then communicates flight plans to individual aircraft. Operators, supervisors, and commanders do this using "volume stacking", where they adjust the volume of all incoming radio channels so the most important are the loudest and the least important are quieter. The aerial supervisor is positioned in an aircraft at the top of the fire traffic area (FTA) in order to see all aircraft below and monitor the fire status. Given the ad-hoc nature of these processes, there is an opportunity to leverage new technologies to upgrade the current system for a more effective response.

B. Scalable Traffic for Emergency Response Operations (STEReO)

The STEReO project investigates the application of NASA UAS and UTM technology to emergency response situations, including wildfire response, to improve operational effectiveness by leveraging vehicle autonomy, improved communications, and software for better coordination [15]. Since currently little data is exchanged between pilots and operators without going through the coordinator, UTM has the potential to reduce the load of the aerial supervisor. Additionally, UAS could be used as relays for these communications to increase communication bandwidth and resilience. While currently UAS are sometimes used in wildfire response, their use is limited to collecting data and providing situational awareness. Sections of the airspace must be cordoned off to allow for UAS flight, and therefore UAS do not coordinate fully with other assets. A UTM system that allows UAS to fly in conjunction with other aircraft could thus increase the effectiveness of existing UAS capabilities. This may enable consideration of additional UAS capabilities if integrated properly, including small item delivery and aerial ignition. STEReO has the potential to reduce the load on manned aircraft and reduce the overall risk to human lives [1, 15].

C. Existing Models

There are multiple existing models for wildfire propagation. One of the most influential models was created by Rothermel in 1972, where energy conservation equations were utilized to develop an elaborate fire growth model. Rothermel considered the fire as a sequence of ignitions that could be represented by equations, using only variables that are able to be measured and quantified in real life [6]. Many modern models incorporate Rothermel's equations, including BehavePlus [7], FARSITE [8], and DEVS-FIRE [9]. These models are used by fire responders in some scenarios to determine their best plan of attack. The models include the effects of wind, terrain, fuel type, and humidity, and have been verified a number of times [9, 16, 17]. In addition to modeling fire behavior, other models include effects of suppression efforts [10, 11].

Deng and Liu quantify the benefits of UAS by looking at the load they can carry and therefore reduce from other aircraft [1]. Their research uses back propagation neural networks to evaluate the efficiency of the UAS [1], but their model does not incorporate fire propagation or the other actors in fire response. SMARt-STEReO includes both fire propagation behaviors and suppression efforts, as well as the effect of numerous faults and mitigation efforts.As a result, unlike existing previously-developed stand-alone models, the SMARt-STEReO integrated model can be used to understand both the performance and resilience of wildfire response operations.

D. Resilience

Resilience is, broadly, a system's ability to prevent and mitigate faults, which can include reliability, robustness, recovery, and reconfigurability [18]. The major difference between system resilience approaches and traditional reliability approaches is that resilience approaches consider the adaptation and recovery of the system post-event, rather than only fault prevention [19, 20]. As a result, to assess the resilience of a system, it is important to model the system's dynamic behavior post-event to determine how the system recovers and adapts [21]. Because of this dynamic property, a number of modeling frameworks have been presented to simulate the effect of faults (and the resulting recovery and/or adaptation) to assess resilience, including OpenCossan [22] and a number of libraries for use with Matlab, Simulink, and Modelica models (see [12] for a detailed discussion). While these approaches are useful for simulating resilience in detailed pre-existing models developed in proprietary toolchains, they limit the ability to extend the modeling paradigms to suit fault propagation, visualization, and resilience assessment in more complex problems, where it is often necessary to define a more complex simulation and/or propagation framework. Thus, prior work has developed the fmdtools toolkit to enable a more expressive and adaptable code-based model representation while providing integrated tools for high-level resilience assessment and visualization [12].

This paper uses the fmdtools framework to develop and assess the resilience of the high-level SMARt-STEReO model of wildfire propagation and emergency response. Models in fmdtools are made of two main classes: functions and flows. Functions can have dynamic behaviors, fault modes, variable states, and components, and can represent the agents, components, or tasks that the system performs. Flows, on the other hand, represent the energy, material, signal, or other attribute passed between or shared by individual functions. To simulate a model constructed in this way, the behavior of each function is run and propagated through the model over a set of discrete time-steps. To assess resilience, these simulations are run over a set of fault modes over a set of possible injection times to quantify the statistical expectation of the integral of performance losses for those modes over each simulation [23]. This expectation of performance loss captures the important attributes of resilience (i.e., reliability, robustness, recovery, etc.) so that designs with different hazard responses can be compared. In this paper, the fmdtools simulator is used to assess resilience by simulating individual fault scenarios over a set of different maps and quantifying the resulting performance losses.

III. Model Development

The SMARt-STEReO wildfire response model is built using the python package fmdtools, an open-source python package for simulating the dynamic response of systems to hazardous scenarios and quantifying the associated resilience [12]. As described in Section II, fmdtools models consist of functions, which define model behaviors, and flows, which define connections between functions. In the SMARt-STEReO model, functions constitute the aerial and ground assets (e.g., aircraft, ground crews, incident commander), as well as the propagation of the fire, as shown in Table 1. Flows constitute the communications between aerial and ground assets, as well as the ground and terrain itself. The relationships between these functions and flows are shown in Fig. 1, where nodes with an "H" are helicopters, "T" are air tankers, "GC" are ground crews, "EC" are ground crews with fire engines, "S" are surveillance UAS, and "UAV" are delivery UAS. Using this model, one can compare various wildfire suppression strategies both over sets of nominal operational scenarios (i.e., different fire conditions) and over the hazardous events modeled as faults in Table 3. While a full model description can be found at [24], the next subsections describe the parameters, behaviors, and resilience/performance analysis of this system in further detail to understand the contribution of this paper.

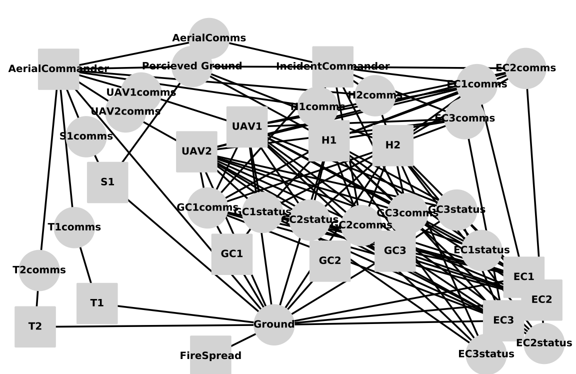


Fig. 1 Network graph illustrating high-level functions and flows in the SMARt-STEReO model.

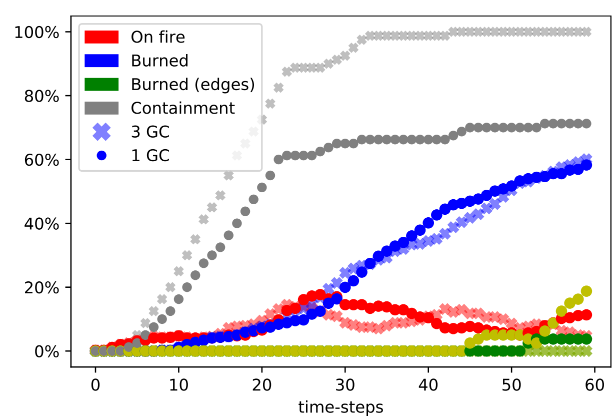


Fig. 2 Fire metric variation due to change in number of ground crews.

Table 1 Model function classes and their respective roles in the model and operational modes.

Function	Purpose	Operational Modes
Incident Commander	Coordinates attack and identifies highest-threat locations; sends locations to aerial supervisor and crews	Identify Threats, Assign Firelines, Move Surveillance Assets
Aerial Supervisor	Updates the perceived fire grid based on surveillance and coordinates drop locations for the tankers and helicopters based on list from incident commander	Move, Ground, Loiter, Refuel
Tanker	Drops retardant or water at points specified by the aerial supervisor	Ground, Refill, Drop, Major Failure
Helicopter	Transports ground crews to their communicated desired locations, either to the fire line or back to base; water drops and surveillance if needed	Surveillance, Drop, Refuel, Delivery, Ground Crew Transport
Ground Crew	Creates fire break at a location specified by the incident commander	New Location, Helicopter Transport, Land Cut, Standby, Rest
Engine Crew	Creates fire lines that are continuous and long	Move, Land Cut, Rest, New Location
UAS	Perform supply deliveries to ground crews when needed	Delivery, To Base, At Base
Surveillance UAS	Performs surveillance	Move, Ground, Loiter, Refuel
Fire Propagation	Models how the fire has changed over a given time step and updates spread four times in one time-step	Fire Spread

Table 2 Model flow classes used to transfer information between functions.

Flow	Purpose	Variables	Connected Functions
Ground	Stores the data for each grid point; allows for the interaction between the fire spread and suppression efforts (drops, firelines)	Time, fuel, fuel type, and altitude for each point	Ground Crews, Engine Crews, Tankers, Helicopters, UAS, Surveillance UAS, Incident commander, Aerial Supervisor, Fire Propagation
Perceived Ground	Used for relaying information about the fire grid to the incident commander in a realistic way, specifically allows for the inclusion of a lag; stores information about where ground crews are located	Time, fuel, fuel type, and altitude for each point	Surveillance UAS, Incident Commander, Aerial Supervisor
Aerial Comms	Used to communicate threat list from incident commander to aerial supervisor	Threat list	Aerial Supervisor, Incident Commander
Asset Comms	Used to communicate fire line locations, drop locations, as well as the current location, future destination, and mode of each asset; each asset has its own communications flow connecting it to different assets	'X' location, 'Y' location, mode, 'X' destination, 'Y' destination, failure	Tanker, Helicopter, Ground Crew, Engine Crew, UAS, Surveillance UAS
Crew Status	Ground and engine crews have additional status updates that need to be communicated besides those in the standard communications; each crew has their own status flow	Level of fatigue, supplies, injury status, efficiency, fire line location, crew location, destination, and operational mode	Engine Crew, Ground Crew

Table 3 Details of the faults and resulting behaviors that can be injected in the model.

Fault	Assets Affected	Behavior
Major Mechanical Fault	Aerial Supervisor, Tanker, Helicopter, UAS, Surveillance UAS	Emergency landing at closest point on the grid that does not currently have a fire, has no fuel, and is not an obstacle in the initial landscape.
Minor Mechanical Fault	Aerial Supervisor, Tanker, Helicopter, Surveillance UAS	Pause any task they were engaged in at the time of injection and return to the airbase. Once they arrive at the airbase, the fault is removed and nominal operation resumes.
Comms Loss	Incident Commander, Aerial Supervisor, Tankers, Helicopters, Ground Crews, Engine Crews, UAS, Surveillance UAS	Dependent on asset affected. Generally they continue current operational states based on previous information and cannot receive new information, such as a new threat list or new drop location. Removed after one time step.
Critical Tool Breaking	Ground Crew, Engine Crew	Regular ground crews go into 'standby' mode until a helicopter or UAS makes a delivery. Once the supply delivery is made, nominal functioning resumes. Engine ground crews travel to the ground base to pick up supplies, resulting in fault removal and return to nominal operations.
Degraded Navigation	UAS, Surveillance UAS	Navigation system is malfunctioning temporarily. The UAS safely lands at the closest safe spot, the fault is removed in one time step, and normal operation resumes.
Tracking Position	UAS, Surveillance UAS	UAS's position is no longer being tracked or communicated. The UAS returns to the base, the fault is removed in one-time step, then normal operation resumes.

A. Parameters and Inputs

Parameters in the model can be adjusted to specify a particular scenario. The grid size and resolution, for example, can be adjusted to suit a variety of map sizes, as well as location and dimensions of cities, air bases, ground bases, and water sources. In Figure 3b, the city is located on the far right of the grid along side the ground and air bases, and a water source is located towards the bottom left, slightly off center. The fire is affected by parameters for the initial spark location, wind direction, wind speed, and whether the grid point attributes are random, uniform, or some other defined pattern. This pattern is used to assign characteristics that govern fire spread, including how long a point takes to catch on fire and how long it burns. A uniform pattern sets all points to the same values, the random pattern sets each value to a random number within a desired range, and a custom pattern could be created to model a specific scenario. Obstacles are points that have no fuel and can never catch fire, such as a paved road, and they are also created in accordance to the designated pattern.

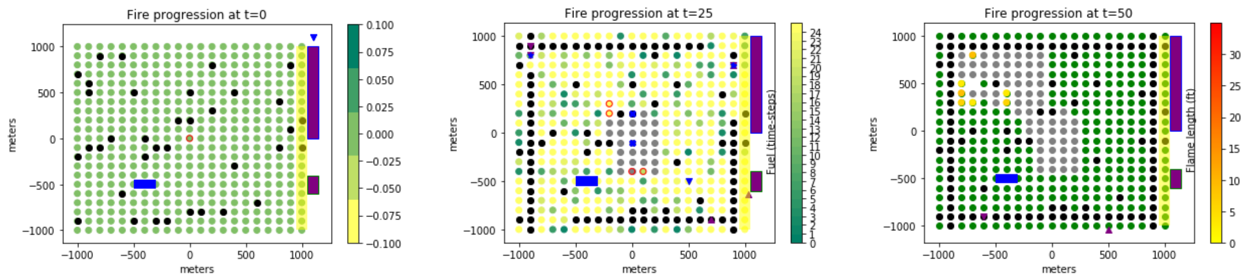
The quantity of each asset can also be adjusted, as can the tanker drop size, efficiency, and quantity. The degree of state awareness of the response team regarding the fire location and spread can be set as either "perfect" or "perceived", with the latter more accurately representing real operators who rely on information gained from surveillance missions for decision-making. Fire lines are formed in a perimeter around the fire, where each side is located at a given length from the origin of the grid. The protection of different sides can be prioritized using different modes. Altitude and spacing for all aerial assets can also be specified. Air tankers, helicopters, UAS, and the aerial supervisor all tend to fly at different altitude ranges, so this parameter allows the user to visualize how the airspace is used [14].

The model grid size, resolution, time step, and asset speeds are all interconnected and must be appropriately tuned for a given scenario. In this research, a 2000m by 2000m grid with an eight-minute time step is selected. The model is run for sixty timesteps, which corresponds to a total response time of eight hours. The timestep is selected based on the desired grid size while considering ground crew and aerial asset speeds. Other important variables, such as aircraft drop impacts, crew fatigue, and tool resupply can be adjusted to appropriately match the timestep.

B. Outputs and Performance Analysis

To understand the results of simulations, a number of different visualizations have been developed which display the dynamic behaviors and the effects of different responses. An integrated animation of the grid displays where drops are conducted, the movement of assets, the formation of fire lines, and the propagation of the fire can also be displayed if desired. Behavior in response to various faults can be displayed, as seen in Fig. 4a and 4b.

Various performance metrics are used to determine how successfully a response contains the fire. Some metrics the model calculates are the total acres burned, the percent of the grid that is burned, the percent that is currently on fire, the



(a) Altitude per grid point. This simulation uses a uniform altitude map.

(b) Fuel per grid point. Black points indicate no fuel. Grey points indicate all fuel is burnt.

(c) Flame length of the grid points currently on fire.

Fig. 3 Altitude, fuel, and flame length at progressive simulation times.

percent of the edges of the grid that are burned, and the percent of the edges of the grid that are on fire, seen in Fig. 2. The total area provides insight on the response's ability to minimize spread, while the edges gives understanding on the response's ability to contain the fire. Burned metrics are useful for analyzing the total destruction from an entire simulation and fire metrics can show when the fire itself is growing or decreasing in size. Another metric used to determine success examines whether an area of the grid designated as a city catches on fire, and if so what percentage of it is affected. This metric helps determine the response's ability to protect valuable areas and assets. In Fig. 2, it is evident three ground crews results in a better response than just one ground crew.

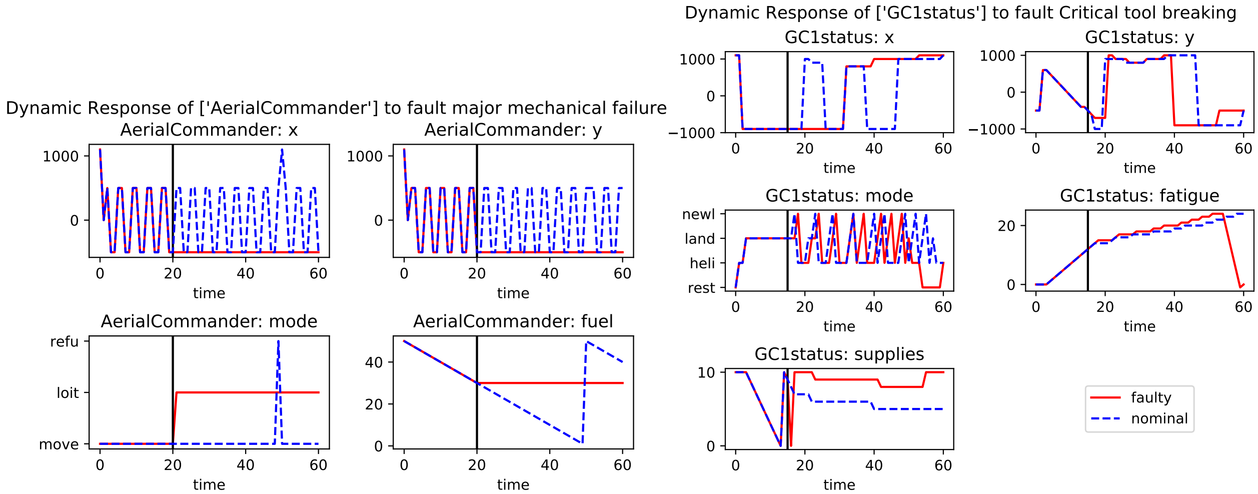
C. Fire Propagation and Suppression Response

The fire propagation function models how the fire changes over a given timestep.Each point in the grid has four properties of interest to the fire propagation: *time*, *fuel*, *fuel type*, and *altitude*. Initial values for these properties are assigned based on the pattern described in Section III.A. In this context, time refers to how long the grid point takes to catch on fire, fuel represents the amount of fuel at a point, fuel type impacts how hot the fuel burns, and altitude is used to model the topography of the grid. If a point is adjacent to a point that is on fire, the point's time value is adjusted based on the slope between the grid points and wind. While a point is on fire, it has an additional property – flame length, which is calculated using the fuel type. The point remains on fire until the fuel runs out. Fig. 3b depicts the fuel characteristics of the grid at $t = 25$, while Fig. 3c shows the flame lengths at $t = 50$. There are four points on fire at $t = 25$, as seen by the red outline in Fig. 3b, and seven points at $t = 50$ shown with colored points in Fig. 3c.

The ground information flow stores the data for each grid point.This flow is connected to the wildfire spread function and each of the assets. The wildfire spread function uses the ground information flow to determine how the fire will spread and updates the grid accordingly. The tanker aircraft, helicopters, and ground crew suppression efforts are represented on the grid in the ground information flow. When a tanker makes a drop or a ground crew builds a fire line, it affects the corresponding grid points. In Fig. 3b and 3c the construction of the firelines occurs at $x = -900$, $x = 900$, $y = -900$, and $y = 900$. Suppression efforts are coordinated and performed by the incident commander, aerial supervisor, tanker aircraft, helicopters, ground crews, ground crews with engines, and UAS.

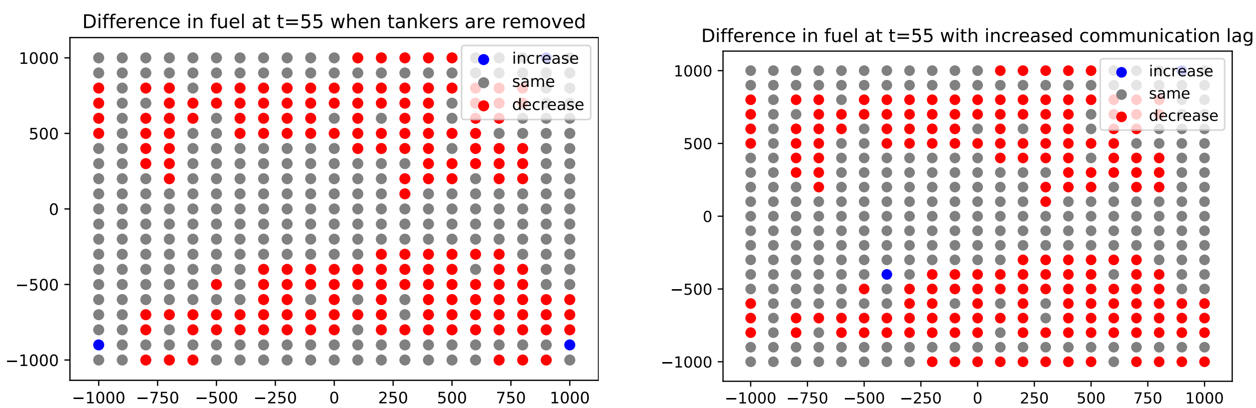
D. Faults

The injection of faults into a system is necessary to understanding the system's resilience overall and in response to a given fault. Fmdtools enables faults to be injected into functions at a given time, in addition to enabling a variety of resilience analyses [12]. A list of faults, their resultant behavior, and the model components into which they can be injected is displayed in Table 3. The two faults most relevant to this paper are the major mechanical failure to the aerial supervisor and a critical tool fault in a ground crew. These are the most relevant since their effects on model behavior will be examined in depth in Section V. When the aerial supervisor experiences this major fault, they immediately land and can no longer perform surveillance, seen in Fig. 4a. However, the aerial supervisor continues to communicate drop locations to tankers and helicopters, and can reassign surveillance tasks to a UAS or helicopter if available. When the minor critical tool fault occurs in the ground crew, the crew shifts to "standby" mode until a delivery is made by a UAS or helicopter, visible in Fig. 4b.



(a) Major fault. **(b) Minor fault.**

Fig. 4 Dynamic behavior in response to a fault.



(a) Removing air tankers results in more fuel burned. **(b) Communication lag results in more fuel burned.**

Fig. 5 Performance differences due to changes in the response.

IV. Verification and Validation

The SMARt-STEReO model is complex and can only provide insight for experiments that are within its scope, and thus the scope must be appropriately validated and verified (Q1). Verification is the process of ensuring a model is correctly created with no apparent bugs or errors in the software or model itself. On the other hand, validation ensures the model actually measures and represents the intended phenomenon. When considering the validity of a model it is important to consider the face validity, testing, and comparison to real data [25]. Face validity for both individual assets and the overall model was partially confirmed through expert validation and comparison to a concept of operations (ConOps) developed for this project, which includes detailed state machines [26]. Preliminary testing is performed to check behavior of the fire propagation as well as individual assets and crews. Individual aspects of the model are also compared to real data. Throughout model development, standard debugging and manual testing practices are followed to verify accurate model behavior. At this stage in development, the model is designed to simulate high level system behavior, and verification and validation efforts are ongoing.

The verification of the behavior of assets is performed in numerous methods. Fig. 5a and 5b are delta graphs that examine the impact of additional crews and decreased situational awareness, respectively. In this instance, the response is analyzed in terms of the amount of fuel at a given grid point. The delta graphs show that, as expected, additional tankers greatly improve the response, while decreased situational awareness degrades the response. Next, the effect of individual assets and crews is verified. First, a baseline response consisting of two air tankers, two helicopters, three ground crews, three engine crews, and no UAS is simulated over 100 random fuel grids with the time of fire line completion recorded over each iteration. Then, each asset or crew is incremented separately and the simulation is run over the same 100 grids. Fig. 7 illustrates the results of this testing over 100 random grids. The addition of supplemental crews successfully decreases the fire spread, as evident by the negative relationship between number of assets and time of fire line completion. It is evident that as expected different assets have varying impacts on fire line time completion. Ground crews and engine crews who are directly constructing the lines have the greatest impact on the time of completion, whereas assets with a more indirect contribution, such as tankers and helicopters, have a smaller impact on completion time. Verification is also performed using other performance metrics such as acres burned. These additional verification steps provide similar results, with increases in individual assets resulting in fewer acres burned overall, yet in this case aerial assets (tankers, helicopters, surveillance UAS) have the largest impact. These results support the notion that each type of asset has a tangible effect on the fire suppression efforts, as expected. While these results are sufficient for the purpose of studying our high level experimental questions, they do not indicate external validity or generalizability at a granular level, and thus it would be inappropriate to assume the impacts of each asset are identical to their real life counterparts without further analysis and development.

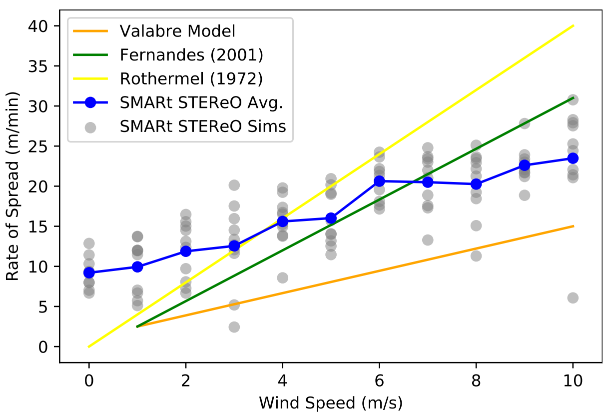


Fig. 6 Comparison of effect of windspeed on fire propagation in SMARt-STEReO to other models.

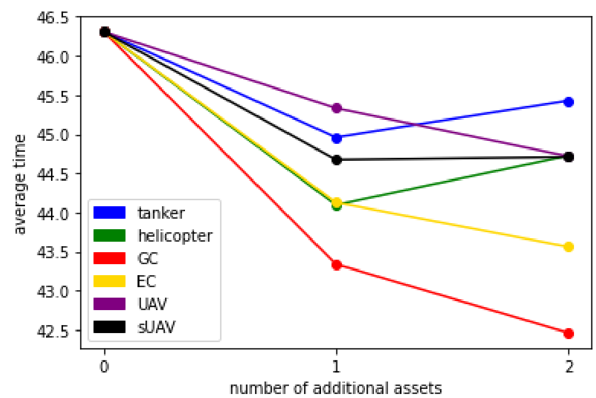


Fig. 7 Asset verification by showing influence of assets on fire line completion time.

The model was validated to an acceptable level of fidelity through extensive research on fire propagation and firefighting tactics, in addition to conversations with Joey Mercer, the lead Principal Investigator for STEReO. A forty page concept of operations was developed for SMARt-STEReO highlighting current firefighting procedure using state-machines, fire propagation trends, and the attempts made to match SMARt-STEReO to the research [26]. This document highlights the overall structure of fire response, the roles of each actor, the speeds at which actions occur, the communication structure, and the responses to failures. Once the model was established, a face validity check was performed by presenting a comprehensive demonstration of the model to Mercer. Important changes to the communication protocol were implemented following Mercer's feedback. Overall Mercer's feedback supports the preliminary face validity of the model structure, seen in Fig 1; however, additional validation efforts are necessary.

Further validation efforts are performed on the most important factors that impact fire growth, that is wind, terrain, and fuel type. First, the effect of wind on the fire propagation model is validated. A control test is run on a uniform grid without intervention or wind to confirm that the baseline rate of spread is acceptable. Next, using a random gird pattern, the wind speed is incrementally increased by 1 m/s, recording the average rate of spread, until the wind speed reaches 10 m/s. The average rate of spread over these trials is compared to existing fire propagation models to validate that the wind trends are realistic. Fig. 6 compares the fire propagation to various other models and confirms that the model generally follows the expected trends for rate of fire propagation relative to wind speed under approximately 10 m/s; however, when the wind speed is low, its accuracy is decreased. While this is not a perfect model for wind's effect on fire spread, these results are sufficient for the purpose of preliminary experiments regarding the high level benefits of implementing STEReO. An additional test is performed to confirm that the model accurately accounts for terrain. A 30° uphill slope has the ability to double the rate of spread [13]. Downhill slope generally lowers the rate of spread [13], so it is assumed that a 30° downhill slope would half the rate of spread. Both 30° uphill and downhill slopes are tested with a uniform grid pattern. The baseline rate of spread with no influence from terrain, wind, or grid pattern variation is 6.25 m/min. From the uphill test, the rate of spread is measured at five different time increments and found to be 12.5 m/min at all of them, or exactly double the baseline rate of spread. In the downhill test, the rate of spread is again measured at five time increments and is found to be roughly 4 m/min. This does not meet the goal of half of the baseline value, but it is the lowest rate of spread allowed by the model without the fire extinguishing itself. As this value is close to the goal value, and because the goal value was chosen somewhat arbitrarily to begin with, this is deemed the highest level of validation possible given the current research and acceptable for the desired resilience analysis. The model incorporates factors for the fuel type and fuel amount that are currently unit-less, but in the future this can be modified to match with moisture content data gathered from actual fires. Currently, the values allow for accurate fire propagation with respect to wind and terrain, so the current iteration is sufficient until more detailed investigation can take place.

SMARt-STEReO is only valid for the scope for which it is designed. The model is designed for macro-level analysis of the resilience and effectiveness of fire response for different response types, specifically for comparing responses with and without UAS. It is not designed to be a comprehensive micro-level fire prediction model, but instead a model of fire propagation and response with a level of accuracy that allows for an understanding of the high-level impact of faults and new technology on the fire response. Fires under high winds, very low winds, or in extreme terrain are less accurately modeled in this stage of development. All ground crews and engine crews are equally effective. However, in a real-world response, some crews may be larger and thus more effective than others. Further, the grid size significantly affects the granularity of the model and, in this research, the smallest size for any object in the grid is 100m x 100m. While the grid size and grid resolution can be changed, this requires re-scaling many parameters, including flight and ground crew speeds. This model has limited sensitivity to various factors, including the physics in the take-off and landing of aerial assets, and weather conditions that may impact aerial assets, such as visibility and heat. An additional limitation of the model is that the quantity of each asset is fixed for the duration of the simulation, whereas in reality, additional assets are added or removed depending on how the fire responds to intervention efforts.

Specific components of the model, including wind speed, terrain, state-awareness, and the addition of various assets, are verified to behave within reasonable range of similar models and/or existing descriptions of behavior. Additionally, the overall model is compared to real data and expert experience. Specifically, the rate of injury sustained to ground crews, the speeds of all aircraft, and fireline building speed have all been compared to real data. More aspects of the model, such as complex terrain and the effect of specific assets like helicopters, need to be critically compared to real data to increase external validity, or the ability of the model to accurately portray a wide range of real circumstances. Ultimately, the model performs as expected over a large range of input parameters and each aspect of the model is evaluated for accurate behavior within the scope of our investigation.

V. Experimental Analysis

A. Methodology

Monte Carlo simulation is used to test whether there is a difference in performance when UAS is added into wildfire response (Q2a), whether the difference is apparent in both nominal and off-nominal cases (Q2b), whether the addition of UAS in wildfire response increases resilience in the system (Q3a), and at what effect size (Q3b). A baseline model without UAS and a version with UAS are tested. The nominal condition is tested by simulating the models with no faults injected. For the off-nominal condition, the effect of two faults are examined individually, one major and one minor. In both models, the major fault is a major mechanical failure to the aerial supervisor, which is injected at $t = 20$. The minor fault is separately injected at $t = 20$ and is a ground crew critical supply fault. The major mechanical fault to the aerial supervisor results in emergency landing and discontinuing nominal behavior for the remainder of the simulation, while the ground crew supply fault results in a standby mode until a delivery is made.

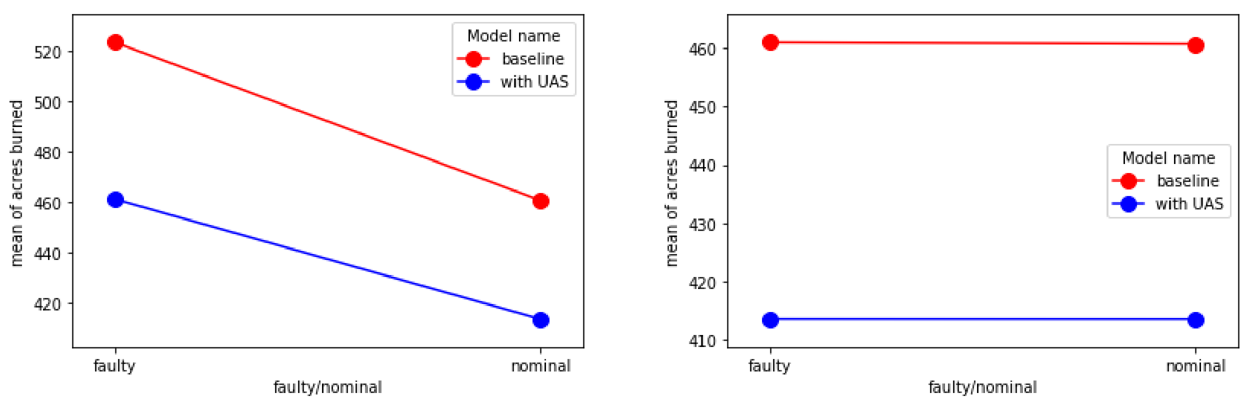
Monte Carlo simulations are useful for uncertainty quantification, and the uncertainty in this case is the average performance and resilience of a model of interest. The random component for the simulation is the initial grid's fuel distribution. For each iteration of the simulation, a new random grid is generated, then each model variation is simulated for 60 time steps over that grid. The simulation is run 2000 times for each condition, resulting in 2000 different grids. All conditions are executed on a 2000m x 2000m grid with resolution 100m x 100m. The water source, air base, ground base, and city locations are constant across conditions and grids. For simplicity, the wind speed and direction are set to 0 across all conditions. Other constants include the fire line placement, which is set to 900m from the ignition source on all sides, the drop size and efficiency of tankers, and ground crew status variables (maximum fatigue, maximum supplies, rest per time step). The amount of tankers, helicopters, ground crews, and engine crews are held constant. The only difference between the baseline and UAS model is the addition of two delivery UASs and one surveillance UAS.

Table 4 Table of performance metric means with standard deviations ($\mu \pm \sigma$).

faulty/nominal	model name	total % burned	total % fire	edges % burned	edges % fire	acres burned
major fault	baseline	52.546 ± 18.870	2.660 ± 3.537	12.442 ± 13.128	1.668 ± 3.064	523.574 ± 188.780
major fault	with UAS	46.436 ± 19.676	2.686 ± 3.257	7.542 ± 10.624	1.053 ± 2.283	491.185 ± 198.182
minor fault	baseline	46.264 ± 20.883	1.798 ± 2.365	7.339 ± 13.046	0.695 ± 1.714	461.013 ± 208.713
minor fault	with UAS	41.637 ± 20.822	1.859 ± 2.186	4.883 ± 10.313	0.616 ± 1.650	413.537 ± 207.449
nominal	baseline	46.236 ± 20.820	1.757 ± 2.320	7.113 ± 12.801	0.662 ± 1.651	460.744 ± 208.091
minor fault	with UAS	41.633 ± 20.712	1.869 ± 2.217	4.587 ± 10.099	0.533 ± 1.504	413.519 ± 206.396

B. Results: Q2

Research question 2 investigates whether there is a difference in performance when UAS is added to the system. If there is a difference, we examine if it is apparent in both nominal and off-nominal conditions. We predict that the implementation of UAS results in better performance when compared to the baseline model (H2a), and this difference will be equally, if not more, apparent in off-nominal circumstances (H2b). To answer both research questions, a 2 (baseline vs with UAS) X 2 (nominal vs faulty) repeated measures ANOVA is conducted on the data from Monte Carlo simulation. A main effect of implementation of UAS on performance is examined to answer question 2a. To answer the second research question, an interaction between operational circumstances and implementation of UAS is examined. As expected, injecting faults results in decreased performance across all conditions. The means and standard deviations for each performance metric are in Table 4. From the table it appears that in all nominal, major fault, and minor fault scenarios the UAS implementation results in better performance. This is evident by the UAS operational condition having a lower average total percentage of the grid burned, grid edges burned, grid edges on fire, and acres burned. As seen in Table 4, the operational conditions have similar standard deviations for each performance metric.



(a) Major fault. **(b) Minor fault.**

Fig. 8 Change in acres burned in response to a fault [27].

Fig. 8a displays the preliminary results of the first 2 (baseline vs UAS model) x 2 (faulty vs nominal) repeated measures ANOVA. The metric of interest is acres burned. As hypothesized, a main effect of model type is evident. From this visualization, it appears that on average the implementation of UAS results in fewer acres burned when compared to the baseline model. Additionally, the injection of a fault results in more acres burned, on average, across model types. Fig. 8a does not clearly illustrate an interaction between the variables, and it appears the difference between the nominal and faulty performances are similar for both the baseline and UAS models.

When conducting an analysis of variance, it is important to examine the spread of data for each condition and identify whether the variance is approximately equal across conditions. From Table 4 the standard deviation across conditions is in a small range of 10. Box plots visually display the spread of the data and are used to further confirm that all experimental conditions have ranges of data that are reasonable. As our sample size is $n = 2000$, we are assuming normalcy, despite box plots illustrating a very slight left skew in the data. There is debate about the robustness of the ANOVA [28, 29] and research has found it is still fairly accurate if sample sizes are large when data does not follow a perfect normal distribution [29–31].

Table 5 2-way Repeated Measures Anova Table for major fault injection

Source	SS	ddof1	ddof2	MS	F	p-uncorrected	p-GG-Corrected	$\eta p2$	eps
faulty/nominal	6.104633e+06	1	1999	6.104633e+06	1538.71018	<0.0001	<0.0001	0.43495	1.0
Model name	6.007546e+06	1	1999	6.007546e+06	170.77928	<0.0001	<0.0001	0.07871	1.0
faulty/nominal * Model name	1.149862e+05	1	1999	1.149862e+05	39.88455	<0.0001	<0.0001	0.01956	1.0

Table 5 displays the resulting test statistic and p-values for the experiment. As seen in the interaction plot, there is a statistically significant difference in acres burned between the baseline and UAS models ($F = 170.779$, $p < 0.0001$), the faulty and nominal operational conditions ($F = 1538.710$, $p < 0.0001$), as well as the interaction between the model type and operation conditions ($F = 39.885$, $p < 0.0001$). Thus, on average, significantly fewer acres are burned when UAS is added to the system, and significantly more acres are burned when a major fault is added to the system. The interaction implies that the difference in performance between the baseline and UAS models is present regardless of nominal or faulty operations. The $\eta p2$ column of the table is the value for the partial eta effect size of each difference. Whether a model is operating in faulty or nominal conditions has the greatest effect on acres burned, with a large effect size of $\eta p2 = 0.43495$. The implementation of UAS into the system has a lesser effect on acres burned with effect size $\eta p2 = 0.07871$. The interaction of the two independent variables has the smallest effect, with a small effect size of $\eta p2 = 0.01956$.

Table 6 Pairwise t-tests table for major fault injection post-hoc testing

Contrast	faulty/nominal	A	B	Paired	Parametric	T	dof	Tail	p-unc	BF10	cohen
faulty/nominal	-	nominal	faulty	True	True	-39.22640	1999.0	two-sided	<0.0001	8.073e+245	-0.31639
Model name	-	with UAS	baseline	True	True	-13.06825	1999.0	two-sided	<0.0001	8.632e+33	-0.28007
faulty/nominal * Model name	nominal	with UAS	baseline	True	True	-10.83720	1999.0	two-sided	<0.0001	1.405e+23	-0.22787
faulty/nominal * Model name	faulty	with UAS	baseline	True	True	-14.28619	1999.0	two-sided	<0.0001	3.481e+40	-0.32407

Table 6 displays the results of the pairwise t-tests, which further confirms the nature of the differences observed in the initial table. On average, significantly more acres are burned in the faulty condition compared to the nominal ($T = -39.226$, $p < 0.0001$). The implementation of UAS results in significantly fewer acres burned on average when compared to the baseline model (T=-13.068, p<0.0001). Under nominal conditions, implementing UAS still results in fewer acres burned on average ($T = -14.286$, $p < 0.0001$). The significant difference in acres burned between the baseline and UAS models is also present under faulty conditions ($T = -14.282$, $p < 0.0001$). Operating in nominal versus fault conditions has a moderate effect on acres burned, with $d = -0.316$. The addition of UAS also has a moderate effect on acres burned ($d = -0.280$), but it is slightly less than the effect of the operational condition. The effect of implementing UAS drops to approximately $d = -0.228$ when operating only in nominal conditions. Implementing UAS has a stronger effect on acres burned when operating in faulty conditions, with $d = -0.324$. Overall operating in a nominal versus faulty condition, as well as the implementation of UAS both have a moderate effect on the average amount of acres burned; however, the implementation of UAS has the greatest effect when operating under faulty conditions.

Fig. 8b displays the interaction plot for a minor fault. The performance metric of interest in this case is acres burned. The figure displays a less extreme difference between the nominal and faulty modes when compared to the major fault. Fig. 8b does display a difference in performance between the baseline and UAS models, regardless of fault. As hypothesized, once again a main effect of implementation of UAS on performance is seen here. Box plots are examined to determine whether variances across conditions are similar. The data is slightly skewed left, similar to the data for the major fault. The variances have slight differences, specifically in the baseline model; however, the differences are minor, and the use of n=2000 samples allows the ANOVA results to still be useful [29–31].

The 2x2 repeated measures ANOVA for the injection of a minor fault are displayed in Table 7. Unlike the major fault in Table5, there is no statistically significant difference in the average acres burned between the nominal and fault conditions ($F = 0.116$, $p > 0.05$), yet there is a significant difference in average acres burned between the baseline

Table 7 2-way Repeated Measures Anova Table for minor fault injection

Source	SS	ddof1	ddof2	MS	F	p-uncorrected	p-GG-Corrected	$\eta p2$	eps
faulty/nominal	4.108181e+01	1	1999	4.108181e+01	0.11603	0.73341	0.73341	0.00006	1.0
model name	4.484100e+06	1	1999	4.484100e+06	117.69954	<0.0001	<0.0001	0.05561	1.0
faulty/nominal * Model name	3.176391e+01	1	1999	3.176391e+01	0.09526	0.75762	0.75762	0.00005	1

and UAS models ($F = 117.700, p < 0.0001$). There is no statistical significance in the interaction between the two independent variables ($F = 0.095, p > 0.05$). The lack of a significant difference in average acres burned between nominal and faulty operational modes confirms that the injection of this minor fault has less of an effect than the major fault, as also seen in the interaction graph. The injection of a minor fault has no significant effect on the acres burned, whereas the major fault had a large effect on acres burned. The addition of UAS has a small to moderate effect size of approximately $\eta p2 = 0.056$ on total acres burned.

Table 8 Pairwise t-tests table for minor fault injection post-hoc testing

Contrast	faulty/nominal	A	B	Paired	Parametric	T	dof	Tail	p-unc	BF10	cohen
faulty/nominal	-	nominal	faulty	True	True	-0.34064	1999.0	two-sided	0.73341	0.027	-0.00078
Model name	-	with UAS	baseline	True	True	-10.84894	1999.0	two-sided	< 0.0001	1.584e+23	-0.22847
faulty/nominal * Model name	nominal	with UAS	baseline	True	True	-10.83720	1999.0	two-sided	<0.0001	1.405e+23	-0.22787
faulty/nominal * Model name	faulty	with UAS	baseline	True	True	-10.76714	1999.0	two-sided	<0.0001	6.88e+22	-0.22816

The pairwise t-tests in Table 8 provides more insight on the observations from Table 7. When a minor fault is injected, there is no significant difference in the average acres burned between the nominal and faulty operational conditions ($T = -0.341, p > 0.05$). The implementation of UAS results in significantly fewer acres burned on average when compared to the baseline model ($T = -10.849, p < 0.0001$), regardless of operational condition. Under nominal operations, on average the implementation of UAS still results in significantly fewer acres burned ($T = -10.837, p < 0.0001$). A significant difference in the acres burned between the baseline and UAS model is also apparent under faulty operations ($T = -10.767, p < 0.0001$). Implementing UAS into the system has a small to moderate effect size on the number of acres burned, with $d = -0.228$. The use of UAS has the same effect size when implemented under both nominal and faulty conditions, with $d = -0.228$ approximately. Thus, when a minor fault is injected, the effect implementing UAS has on the number of acres burned is about the same in both nominal and faulty operational conditions.

C. Discussion Q2

We examined whether there is a difference in performance when UAS is added to the system (Q2a), and if there is a difference is it apparent in both nominal and off-nominal conditions (Q2b) in research question 2. The results from both minor and major fault experiments support our hypothesis that the implementation of UAS results in increased performance. The implementation of UAS has a slightly larger effect size on performance under faulty operations in response to the major fault. Implementing UAS has about the same effect on performance in both nominal and faulty operations in response to a minor fault. Together these results support our hypothesis that the implementation of UAS results in increased performance when compared to the baseline, and this increase in performance is also apparent under faulty operations.

Table 9 Table of means and standard deviations for resilience in performance metrics ($\mu \pm \sigma$).

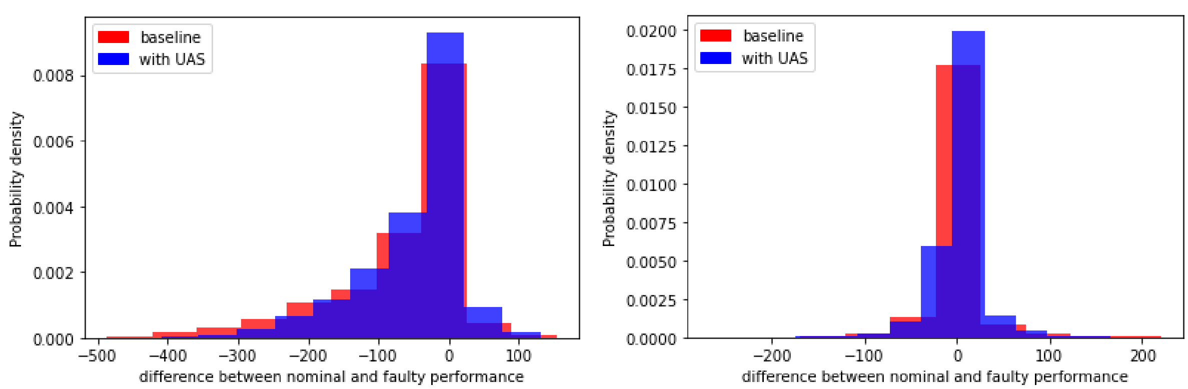
type of fault	model name	total% burned	total% fire	edges % burned	edges % fire	acres burned
major fault	baseline	-6.310 ± 9.110	-0.903 ± 3.241	-5.329 ± 7.833	-1.006 ± 3.294	-62.830 ± 90.801
major fault	with UAS	-4.807 ± 7.450	-0.817 ± 2.973	-2.955 ± 5.946	-0.520 ± 2.406	-47.665 ± 73.864
minor fault	baseline	-0.028 ± 2.743	-0.040 ± 1.393	-0.226 ± 3.093	-0.033 ± 1.309	-0.269 ± 27.327
minor fault	with UAS	-0.004 ± 2.524	0.010 ± 1.459	-0.295 ± 2.939	-0.083 ± 1.204	-0.017 ± 25.063

D. Results Q3

Research question 3 examines if the addition of UAS in wildfire response increases resilience in the system in response to an individual fault, and if so, to what effect size is. We predict that the addition of UAS will increase resilience in the system(H3a), and the effect size will be small to medium depending on the fault injected(H3b). The experimental effect size may not accurately represent the true effect given out modeling methodology and limitations to our model. To investigate this question, resilience metrics are analyzed. For the purpose of this analysis, resilience is defined as the difference in performance between the nominal and off-nominal operational conditions. The average resilience is shown in Equation (1), where $n = 2000$. Based on this calculation, a smaller number or difference between nominal and faulty performance indicates better resilience. Resilience is defined by (1) to make the interpretation and implications clear in the context of a real fire. Because this calculation is looking at the difference between nominal and faulty in an individual model, different models could have the same resilience despite one model performing better than the other on average across faulty and nominal circumstances. This metric is calculated for each condition, the baseline model and the model with the implementation of UAS, using the data from the same Monte Carlo simulation.

$$average\ resilience = \frac{\sum_{i=1}^{n}(nominal_i - faulty_i)}{n} \tag{1}$$

The means and standard deviations for the resilience calculated from each performance metric are found in Table 9. A negative mean signifies that a higher amount of a given value occurred in the faulty operational condition, such as more acres burned. From the table, it appears that when a major fault is injected into the system, the implementation of UAS results in a smaller magnitude of the resilience metric across all performance metrics, indicating potentially better resilience. Thus when UAS is implemented, the difference in performance between the nominal and off-nominal operations is less when compared to a model without UAS. The same pattern is also seen for the minor fault, however the means of the baseline and UAS models are much more similar than those of the major fault. In Table 9, the standard deviations for the baseline and UAS models are within a reasonable range, but they are not precisely equal. For the purpose of this analysis this is permissible, but results should be considered with some caution.The statistical tests performed are designed for use on data with a normal distribution and equal variance, and thus the true difference in results may be lesser or greater than the test suggests. A paired-samples t-test, along with effect size calculation, is performed to examine if these initially observed differences in resilience between the models with and without UAS are significant. The metric of interest is resilience in acres burned, thus the difference in acres burned between the nominal and the faulty scenarios.



(a) Major fault. **(b) Minor fault.**

Fig. 9 Resilience as measured in terms of acres burned in response to faults [27].

When performing a t-test, the data should follow similar distribution shapes. The difference between the sampling means of resilience to a major fault for the baseline and UAS models fits an approximate normal distribution with $R^2 = 0.943$. Fig. 9a shows histograms of the experimental data for a major fault and confirms that the two models, baseline and with UAS, appear to have the same distribution type. From the histograms, it appears that the resilience has a greater spread and could be either slightly smaller or larger for the baseline model. This is evident by the tail of the baseline model extending further to the left. To confirm if there is a significant difference, and if so, what the effect size is, the results from the t-test must be considered.

Table 10 T-test Table for resilience in performance for major fault injection

T	dof	tail	p-val	CI95%	cohen-d	BF10	power
-6.3154	1999	two-sided	<0.0001	[-19.87, -10.46]	0.1832	9.215e+6	1.0

Table 10 displays the results of the t-test performed comparing the difference in performance in response to a major fault between the baseline and UAS models. The test confirms there is a significant ($T = -6.315$, $p < 0.0001$) difference in resilience between the two models. On average, the difference in acres burned between the baseline and UAS models is between -19.87 and -10.46 in approximately 95% of samples. So given a major fault, it is expected that the baseline model will experience a difference between nominal and faulty performance between 19.87 and 10.46 more acres burned than the difference between nominal and faulty performance in the UAS model. The Cohen's D effect size is moderate with $d = 0.183$. These results indicate that the implementation of UAS has a moderate effect on resilience in performance during a major fault.

The same analysis was conducted for the minor fault, beginning with a graph of the distribution of differences in the sampling mean between the baseline and UAS models. The performance metric of interest for the minor fault is again resilience. The differences in the sampling mean do not fit the normal distribution as well as the distribution for the major fault. In this case $R^2 = 0.8457$, which is a moderate fit. Due to the non-normal distribution, the analysis in this case may be prone to errors, but the high sample size of $n = 2000$ makes the analysis robust [29–31].

From examining the histogram of the resilience data in Fig. 9b prior to the t-test results, it is evident the distributions of resilience for the baseline and UAS models look almost identical. There are no obvious differences in the distributions, so in our model the UAS implementation may not have a significant effect on resilience in response to a minor fault. These histograms also visually show that the two distributions are similar in shape and spread.

Table 11 T-test Table for resilience in performance for minor fault injection

T	dof	tail	p-val	CI95%	cohen-d	BF10	power
-0.308647	1999	two-sided	0.7576	[-1.85, 1.35]	0.0096	0.026	0.07141

The results of the t-test in Table 11 confirm the results shown in the histogram. When resilience to a minor fault is measured as the change in the amount of acres burned between nominal and faulty operations, the implementation of UAS does not result in a significant difference in resilience when compared to the baseline model ($T = -0.309$, $p > 0.05$).

E. Discussion Q3

In research question Q3 we sought to answer if the addition of UAS in wildfire response increases performance resilience in the system in response to an individual fault(Q3a), and if so, to what effect size(Q3b). We hypothesized that implementing UAS will increase resilience in the system (H3a), and the effect size will be small to medium depending on the fault injected (H3b). The results from the injection of the major fault support the hypothesis 3a that UAS increases resilience in response to a fault, however, the results from injection of a minor fault do not fully support this hypothesis. There was no significant difference in the resilience metric between the UAS and baseline models, yet the UAS model is still preferred as it performs better than baseline across faulty and nominal operations. Implementing UAS had a moderate effect size on performance resilience when a major fault was injected, while there was no significant effect when the minor fault was injected. These results support our hypothesis 3b that the effect size depends on the fault injected. The model has certain limitations, including limited sensitivity to small changes, such as a minor fault. In a more sensitive model, the injection of a minor fault could result in different results.

VI. Conclusion

The SMART-STEReO project demonstrates the safety and performance benefits of STEReO through system modeling and analysis of resilience. SMART-STEReO accomplishes its goals by a dynamic simulation of the system's performance in nominal and faulty scenarios. Initial verification efforts have been conducted on the model, including debugging procedures and analyzing results from specific inputs. Initial validation efforts are completed through expert consultation and comparisons to existing models. Ultimately the model has some limiting assumptions, but it is verified and validated sufficiently for the purpose of analyzing resilience in the wildfire response system at a high level (Q1). The model is used to experimentally test whether the implementation of STEReO improves performance (Q2a), and if this improvement is also found in off-nominal operational conditions (Q2b). Results from the Monte Carlo simulation support the hypotheses that the implementation of STEReO results in increased performance in terms of fewer acres burned (H2a) and in off-nominal circumstances, the increase in performance is equally, if not more, apparent (H2b). We also test if the implementation of STEReO resulted in more resilience in performance (Q3a), and if so, to what effect size (Q3b). Results indicate that when a major fault is injected, there is a statistically significant difference in resilience between the baseline and UAS models and the implementation of STEReO has a small to moderate effect on resilience. When a minor fault is injected, there is no significant difference in resilience between the baseline and UAS models. Overall the results support the hypothesis that the implementation of STEReO results in increased resilience (H3a), as well as the hypothesis that the effect size is small to medium but dependent on the severity of the fault injected (H3b).

Future iterations will allow for a more versatile model, which will increase the scope of possible analyses. More extreme variations in terrain and weather conditions, such as the presence of smoke, will be considered. For the purpose of this paper, resilience was quantified as change in performance between nominal and faulty operations. However, considering risk is equally, if not more so, important. The quantification of risk associated with various faults needs to be analyzed more thoroughly in the future. Due to the human involvement and risk and firefighting, it will be beneficial to demonstrate that the implementation of STEReO results in higher resilience due to less human risk. Further, our experimental analysis focused only on two faults, which resulted in simplified system behavior when compared to the real-world operational conditions. In future work, more detail will be added to faults, in addition to more faults overall. Some future analysis to further assess the safety, risk, and resiliency will use adaptive stress testing on the SMARt-STEReO model. This will allow for the identification of the most probable scenarios that result in undesirable outcomes.

Acknowledgments

This research was partially conducted at NASA Ames Research Center. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government.

We would like to thank the following people:

- Joey Mercer, primary investigator for the STEReO project, for taking time to discuss our model and provide feedback.
- Guillaume Brat
- Haley Feck and Abel Morelos for coordinating logistics that allowed for the internship experience.
- University Space Research Association for partially funding this research.

References

[1] Deng, L., He, Y., and Liu, Q., "Research on Application of Fire Uumanned Aerial Vehicles in Emergency Rescue," *2019 9th International Conference on Fire Science and Fire Protection Engineering (ICFSFPE)*, 2019, pp. 1–5. doi: 10.1109/ICFSFPE48751.2019.9055875.

[2] Linkov, I., Bridges, T., Creutzig, F., Decker, J., Fox-Lent, C., Kröger, W., Lambert, J. H., Levermann, A., Montreuil, B., Nathwani, J., et al., "Changing the resilience paradigm," *Nature Climate Change*, Vol. 4, No. 6, 2014, pp. 407–409.

[3] Punzo, G., Tewari, A., Butans, E., Vasile, M., Purvis, A., Mayfield, M., and Varga, L., "Engineering resilient complex systems: the necessary shift toward complexity science," *IEEE Systems Journal*, 2020.

[4] Romich, R., and et al, "Although FAA Has Taken Steps to Improve Its Operational Contingency Plans, Significant Work Remains to Mitigate the Effects of Major System Disruptions," Tech. Rep. AV2017020, Federal Aviation Administration, 2017.

[5] Yodo, N., and Wang, P., "Engineering resilience quantification and system design implications: A literature survey," *Journal of Mechanical Design*, Vol. 138, No. 11, 2016.

[6] Rothermel, R. C., "A mathematical model for predicting fire spread in wildland fuels," , 1972.

[7] Andrews, P. L., Bevins, C. D., and Seli, R. C., "BehavePlus fire modeling system, version 4.0: User's Guide," *Gen. Tech. Rep. RMRS-GTR-106 Revised. Ogden, UT: Department of Agriculture, Forest Service, Rocky Mountain Research Station. 132p.*, Vol. 106, 2005.

[8] Finney, M. A., *FARSITE, Fire Area Simulator–model development and evaluation*, 4, US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 1998.

[9] Gu, F., Hu, X., and Ntaimo, L., "Towards validation of DEVS-FIRE wildfire simulation model," *Proceedings of the 2008 Spring simulation multiconference*, 2008, pp. 355–361.

[10] Seraj, E., Wu, X., and Gombolay, M., "FireCommander: An Interactive, Probabilistic Multi-agent Environment for Joint Perception-Action Tasks," , 2020.

[11] Ntaimo, L., Hu, X., and Sun, Y., "DEVS-FIRE: Towards an Integrated Simulation Environment for Surface Wildfire Spread and Containment," *Simulation*, Vol. 84, 2008, pp. 137–155. doi:10.1177/0037549708094047.

[12] Hulse, D., Walsh, H., Dong, A., Hoyle, C., Tumer, I., Kulkarni, C., and Goebel, K., "fmdtools: A Fault Propagation Toolkit for Resilience Assessment in Early Design," *International Journal of Prognostics and Health Management*, Vol. 12, No. 3, 2021.

[13] of Agriculture, U. D., and Service", U. F., *Safe and Effective Wildfire Response*, (accessed August, 2020). URL `https://www.fs.usda.gov/managing-land/fire/response`.

[14] Groupd, N. W. C., *Wildland Fire Suppression Tactics Reference Guide*, United States Department of the Interior and United States Department of Agriculture and National Association of State Foresters, April 1996. URL `http://www.coloradofirecamp.com/suppression-tactics/suppression-tactics-guide.pdf`.

[15] Mercer, J., Mcswain, R., and Ippolito, C., "Scalable Traffic Management for Emergency Response Operations (STEReO)," *AIAA Aviation Conference*, National Aeronautics and Space Administration, 2020. URL `https://hsi.arc.nasa.gov/groups/AOL/downloads/AIAA-STEReO-2020-v4-web.pdf`.

[16] Grabner, K., Dwyer, J., and Cutter, B., "Validation of BEHAVE fire behavior predictions in oak savannas using five fuel models," 1997.

[17] Arca, B., Duce, P., Pellizzaro, G., Bacciu, V., Salis, M., and Spano, D., "Evaluation of FARSITE simulator in a Mediterranean area," *Proceedings of the 4 th International Wildland fire Conference*, 2007, pp. 14–17.

[18] Shen, L., Cassottana, B., Heinimann, H. R., and Tang, L. C., "Large-scale systems resilience: A survey and unifying framework," *Quality and Reliability Engineering International*, Vol. 36, No. 4, 2020, pp. 1386–1401.

[19] Clark-Ginsberg, A., "What's the Difference between Reliability and Resilience," *Department of Homeland Security. March*, 2016.

[20] Yodo, N., Wang, P., and Rafi, M., "Enabling Resilience of Complex Engineered Systems Using Control Theory," *IEEE Transactions on Reliability*, Vol. 67, 2018, pp. 53–65.

[21] Cassottana, B., Shen, L., and Tang, L. C., "Modeling the recovery process: A key dimension of resilience," *Reliabilty Engineering and System Safety*, Vol. 190, 2019. URL `https://doi.org/10.1016/j.ress.2019.106528`.

[22] Patelli, E., Tolo, S., George-Williams, H., Sadeghi, J., Rocchetta, R., de Angelis, M., and Broggi, M., "OpenCossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis," 2018.

[23] Hulse, D., Hoyle, C., Tumer, I. Y., Goebel, K., and Kulkarni, C., "Temporal Fault Injection Considerations in Resilience Quantification," *Proceedings of the ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Society for Computer Simulation International, St. Louis, MO, USA, 2020.

[24] Hulse, D., Andrade, S. R., Spirakis, E., Walsh, H. S., and Davies, M. D., "SMARt-STEReO: Preliminary Model Description," Tech. Rep. NASA/TM-20205007481, NASA Ames Research Center, 2020.

[25] Carson, J. S., "Verification Validation: Model Verification and Validation," *Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers*, Winter Simulation Conference, 2002, p. 52–58.

[26] Walsh, H. S., Spirakis, E., Andrade, S. R., Hulse, D. E., and Davies, M. D., "SMARt-STEReO: Preliminary Concept of Operations," Tech. Rep. NASA/TM-20205007665, NASA Ames Research Center, 2020.

[27] for Data Science, P., *One-Way ANOVA with Python*, (accessed August, 2020). URL `https://www.pythonfordatascience.org/anova-python/#assumption_check.`

[28] Wells, C., and Hintze, J., "Dealing with assumptions underlying statistical tests," *Psychology in the Schools*, Vol. 44, No. 5, 2007, p. 1026. URL `https://doi.org/10.1002/pits.20241`.

[29] Knief, U., and Forstmeier, W., "Violating the normality assumption may be the lesser of two evils," *bioRxiv*, 2020. doi: 10.1101/498931, URL https://www.biorxiv.org/content/early/2020/05/05/498931.

[30] Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., and Bendayan, R., "Non-normal data: Is ANOVA still a valid option?" *Psicothema*, Vol. 29, No. 4, 2017, pp. 552–557. URL https://www.redalyc.org/articulo.oa?id=727/72753218018.

[31] Troncoso Skidmore, S., and Thompson, B., "Bias and precision of some classical ANOVA effect sizes when assumptions are violated," *Behav Res*, Vol. 45, 2013, pp. 536–546. URL https://doi.org/10.3758/s13428-012-0257-2.