2024 IEEE International Symposium on Safety Security Rescue Robotics (SSRR)
November 12-14, 2024. New York City, USA

Analysis Of Human Errors In Small Uncrewed Aerial Systems Used During Hurricane Ian

Robin R. Murphy[1] and Thomas Manzini[1] and S. Camille Peres[2]

Abstract— **This report documents 51 instances of human error for small uncrewed aerial system (sUAS) for 34 missions flown under the direction of the State of Florida UAS Task Force 1 at Hurricane Ian from 9/29/2022 to 10/2/2022. The current state of human failure analysis for small UAS has been directed at aviation safety, rather than overall mission success. This work focuses on mission success, which includes data product delivery. Using an expansion of an existing failure taxonomy for disaster robotics, the data suggests that human error was not directly responsible any failures of the physical platform, but was the source of four failures to complete missions (terminal failures) and delays in 27 missions in data product delivery (non-terminal failures). Human error occurred in all but three missions. Procedural errors were occurred approximately twice as often as skill-based errors, and most frequently occurred during the Initiation and Termination phases of the mission, not the Execution phase. The documentation and preliminary analysis contributes to robotics and HRI research in interfaces, human error, training, and procedures for small UAS in any domain. The analysis is also expected to have societal impact as it alerts emergency management leadership and UAS pilots to tendencies towards errors during disasters.**

I. Introduction

Fourteen agencies and institutions deployed UAS and pilots, operating under the direction of the State of Florida UAS task force, FL UAS-1, to Hurricane Ian is now considered to be the third costliest hurricane to strike the United States. Hurricane made landfall near Fort Meyers Beach, Florida on 9/28/2022 as a Category 4 storm after briefly attained a Category 5 status. FL UAS-1, directed by Florida State University, deployed nine squads from 9/27/2022 to 10/5/2022, with flying beginning on 9/29/2022 when the hurricane winds had subsided. All pilots, see Fig. 1, were experienced with UAS, had trained or deployed together in regional incidents, and used UAS in the course of their normal work. In the first 4 days of flying, the squads successfully completed 30 out of 34 missions producing 567.6GB of valid mission data. The metrics of mission success are that the data products from the UAS sorties are provided in the time requested. This means that mission success is more than safely or quickly flying over an area, it includes any formatting, post-processing, or labeling of data for distribution to incident command.

However, the missions were not seamless; in addition to the four terminal missions, 27 of the remaining 30 showed evidence of human error, leaving only three error-free missions. The evidence comes from detailed mission logs that were kept for the first four days of flying. These logs provide an opportunity to perform a post hoc analysis of human error, as the logs capture what a drone squad was tasked to do and whether they were able to meet those expectations. The first four days is particularly interesting because missions were all a single type, Recon/Rapid Needs Assessment, and thus can be directly compared. This mission type is the most common for disasters (see [1], [2], [3]) and has high urgency since the resulting data products are used for directing life-saving search and rescue efforts, determining additional resources needed for civilians, and projecting recovery costs.

Fig. 1. Small UAS squads at Hurricane Ian.

An analysis of the logs, summarized in the remainder of the paper, documents that UAS squads were returning with missions with unusable data or data that was in the wrong format or incomplete, requiring unexpected additional work at the forward operating base to produce the data products, delaying their release for dissemination to responders and emergency managers.

II. Related Work

This work is novel in that it explores small UAS pilot error over the different mission phases and its impact on mission objectives, in contrast to papers focusing on aviation safety [4], [5], [6], [7], [8], [9]. It adapts prior work in formal taxonomies for categorizing human error based on previous work in disaster robotics or field robotics, merging the *human* error categories in Carlson and Murphy [10] with Honig and Orad-Gilad [11]. However, the four categories of human error

*This work was supported by the National Science Foundation under Grants No. CMMI-2306453 and IIS-2229881.

[1]Robin R. Murphy and Thomas Manzini are with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, US {robin.r.murphy, tmanzini}@tamu.edu

[2]S. Camille Peres is with the Nuclear Regulatory Commission, Rockville, MD, USA camille.peres@nrc.gov

2024 IEEE International Symposium on Safety Security Rescue Robotics (SSRR) | 979-8-3315-1095-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/SSRR62954.2024.10770053

(mistakes, slips, lapses, and deliberate violations) all require knowledge of the motivation of the human. This paper is limited to error analysis from mission logs; therefore, the squads' motivations are unknown, and thus, the analysis cannot place errors into these categories. However, this paper can associate errors with the type of activity in which the error occurred, either following procedures or protocols (procedural-based activities) or directing the drone (skills-based activities). As shown in Fig. 2, the activity level would be a refinement of the categories.

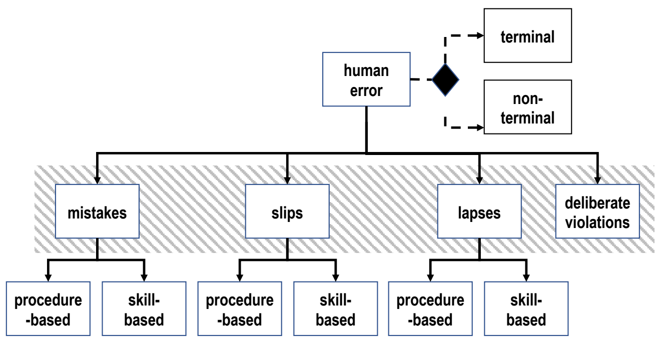

Fig. 2. Graphical representation of the merged human error taxonomy for field robots. Categories of human error that cannot be extracted from mission logs are cross-hatched.

A. Taxonomies of Human Error with small UAS

The literature on human error for small UAS has focused on aviation safety and root causes following the classic human factors analysis and classification system (HFACS) framework for aviation safety [12], which is not helpful for understanding data-to-decision mission success. In a 2020 survey of 69 papers on human factors for UAS [6], only five papers discussed specific sources of UAS pilot error for i) mishaps in general [7], ii) collisions [5], or iii) general aviation safety [4], [8], [9]. None considered the role of the pilot in meeting mission objectives or methods for documenting associated errors. One additional paper discussed the role of pilot error in performing missions but errors were limited to an element within a larger Model-based Systems Engineering framework for human factors analysis [13].

B. Taxonomies of Human Error with Robots in General

Three human-robot error taxonomies [10], [14], [11] are relevant for categorizing and counting the frequency of occurrence of human error in field robots. Fig. 2 represents a novel merger of the specific interaction errors in [10] (mistakes, slips) with [11] (lapses, deliberative violations). However, those interaction errors cannot be determined from mission logs because knowledge of the motivation or intent of the human is required for classification. The taxonomy in Fig. 2 is also novel because it notes that each of the four interaction errors could be further classified by activity, procedure-based, or skill-based, which would help with understanding impacts and root causes. An analysis of human error can capture those activities without understanding the mental state of the human. Other human-robot error taxonomies, notably [15], [16], [17], are outside of the scope of this article as they focus on categorizing failures in terms of impact on the social-emotional aspects of how robots are perceived or how users interact with them socially, which are not germane to field robotics.

Carlson and Murphy [10] put forth a robot failure taxonomy that divided failures into *physical* or *human* and subdivided *human* failures into *mistakes* (performing an action with the wrong intent, e.g., flying a mission using the incorrect technique) or *slips* (attempting to do the right thing unsuccessfully, e.g., accidentally performing a step out of sequence). The taxonomy is mission-oriented, as the impact of each failure is classified as either *terminal*, where the robot failed to complete its mission, or *non-terminal*, where the mission was delayed or degraded but still accomplished in some form. This taxonomy has been used in human-robot interaction research characterizing influences on trust [18].

A similar fault taxonomy was used in [14] for assessing failures in the RoboCup competition entries through 2012. This taxonomy divided failures into hardware, software, algorithms, and interaction, with *interaction failures* being subdivided into human, agent-robot, and environment. However, no data on specific instances or frequency of occurrence of interaction errors were provided. The work did not pose any refinement of interaction errors, in contrast to [10] which subdivided interaction errors into mistakes and slips.

Honig and Orad-Gilad [11] offered a dedicated human-robot failure taxonomy which divided the source of failures into technical failures (software, hardware) and interaction failures (social norm violations, human errors, and environment and other agents). [11] use a similar scheme to [10] for rating the functional severity of human error: *terminal*, *non-critical*, and *recoverable*. Terminal is the equivalent of terminal in [10], while *non-critical* and *recoverable* appears to be subcategories of *non-terminal* in [10], though neither *non-critical* nor *recoverable* were precisely defined in [11]. [11] expands [10] include *lapses*, such as errors in memory or attention (e.g., forgetting to recharge batteries) and *deliberate violations* (e.g., knowingly not following procedures).

III. Human Error Identified at Hurricane Ian

The preliminary analysis of the 51 instances of human error loosely follows Sutcliffe and Rugg's recommendations on what questions to ask in defining root causes of human-machine error [19]. Sec. III-A describes the mission logs and answers the questions of when were the effects of failure apparent? and who observed/reported the failure? Sec. III-B answers what has been observed to have failed? followed by Sec. III-C on what are the consequences of the failures? The question of where did the failure occur? is addressed in two ways, with an examination of what activity did the error occur in (Sec. III-D) and what phase of the mission did it occur in (Sec. III-E).

A. Missions and Mission Logs

In order to follow the analysis, it is helpful for the reader to understand the mission tasking process and how the mission logs were created and maintained at Hurricane Ian. Each morning, the task force leader would receive a list of requests for aerial data from an incident commander, translate those requests into one of the nine types of missions[20], and then allocate one or more missions to squads to cover the requests. The squads were informed of the assignments at the morning briefing and each squad was given a paper form for each mission with the mission name (usually the location where the mission was to be flown) and a form to fill out on sorties. The data managers, co-authors Murphy and Manzini, created a spreadsheet with mission names, mission types, and squads. The form also had a reminder to label all data products using the mission name before giving the data to the data managers.

When the squads returned in the late afternoon, the data managers would log the arrival of the data in the spreadsheet, copy the data, and then begin any mission-specific post-processing (e.g., editing videos, creating orthomosaics). The data managers also attempted to enter into the log any additional processing needed to rectify or mitigate incorrect data products and to make notes as to the effects of the failure and whether it was identified by the squad in the field or by the data managers. In general, most squad members were visibly fatigued and eager to hand off the data so they could get food and start recharging their gear for the next day. Attempts to informally interview the pilots or ask questions beyond casual conversation were met with complaints to the task force leader that the data managers were wasting the pilot's time even though the questions were being asked while the pilots were waiting for the data to be copied. The reluctance of pilots to be interviewed after flying long days during disasters has been documented in other disasters [21].

Human error was detected primarily by the data managers, who could quickly see the returned data was not what had been asked for or was not in the correct format. Some squads self-reported to the data managers that there had been a problem during a flight, while other instances of human error came up in conversation in the evenings or in the informal evening briefing; these were logged opportunistically by the data managers.

B. Overview of Human Error Counts

Table I summarizes the missions by day, the number of missions flown that day, the number of human errors for each mission, and flight style (flown either autonomously or manually). The table shows that there were a total of 34 missions, of which only three had no recorded errors (12, 20, and 27). 20 of the missions were autonomous, and 17 were manual, with missions 8, 13, and 17 involving both autonomous and manual flights. Missions 8 and 13 contained errors in only one of the flight styles, while mission 17 had errors occur in both the autonomous and manual portions of the mission.

The only mission type for the four-day period was the Recon/Rapid Needs Assessment, which had manual and autonomous variants of flight style. The flight style was specified at the time of tasking. One was targeted rapid inspection of infrastructure damage (e.g. bridge inspection); these flights are typically flown manually in first-person view mode. The data product for this variation is still images of significant damage and video snippets. A data manager would typically select several 30-second to 3-minute snippets from the longer 12-20 minute flight videos for dissemination as the final data product. The other was rapid low-resolution mapping of an area of interest, flown autonomously with a software flight and camera control package that collected raw images. The images would be merged into the data product, an orthomosaic map, by the squad (if they had the post-processing software) or a GIS specialist at the forward operating base.

The temporal pattern of mission variants followed the previously observed patterns at Hurricanes Harvey [3], Michael [2], and more recent hurricanes, which suggests that the

TABLE I

OVERVIEW OF MISSIONS AND ERRORS BY DAY. A WHITE CELL INDICATES THE MISSION WAS FLOWN WITH AN AUTONOMOUS OR MANUAL FLIGHT STYLE. MISSIONS 8, 13, AND 17 USED BOTH FLIGHT STYLES.

Date	# Missions	Mission ID	# Errors	Autonomous Flight	Manual Flight
9/29/2022	6	1	2		
		2	2		
		3	1		
		4	2		
		5	2		
		6	1		
9/30/2022	5	7	1		
		8	1		
		9	1		
		10	1		
		11	1		
10/1/2022	14	12	0		
		13	1		
		14	3		
		15	1		
		16	1		
		17	2		
		18	1		
		19	2		
		20	0		
		21	1		
		22	1		
		23	2		
		24	1		
		25	4		
10/2/2022	9	26	3		
		27	0		
		28	2		
		29	3		
		30	1		
		31	3		
		32	1		
		33	1		
		34	2		
Grand Total	**34**		**51**	**20**	**17**

TABLE II

TERMINAL ERRORS COMPARED WITH NON-TERMINAL.

Date	# Missions	# Errors	Non-Terminal Failures (46)		Terminal Failures (5)	
			Autonomous	Manual	Autonomous	Manual
9/29/22	6	10	1	9	0	0
9/30/22	5	5	1	3	1	0
10/1/22	14	20	11	6	1	2
10/2/22	9	16	10	5	1	0
Total	**34**	**51**	**23**	**23**	**3**	**2**

TABLE III

HUMAN ERROR BY FLIGHT STYLE AND ACTIVITY.

			Autonomous (26)		Manual (25)	
Date	# Missions	# Errors	Procedure	Skill	Procedure	Skill
9/29/22	6	10	1	0	4	5
9/30/22	5	5	1	1	0	3
10/1/22	14	20	11	1	5	3
10/2/22	9	16	9	2	4	1
Total	**34**	**51**	**22**	**4**	**13**	**12**

human error is representative for disasters in general. The pattern is for pilots to fly manually for the first days to provide rapid spot checks on Infrastructure Damage that could be used tactically, then begin flying more systematically for Rapid Low-Res Mappiong for documentation.

C. Non-Terminal Compared with Terminal Errors

Table II shows the distribution of non-terminal and terminal errors. The 51 instances of human error had a noticeable impact on mission success, with 27 having avoidable delays or degradations (non-terminal failure), four missions not completed (terminal failure), and only three of the 34 missions being completed without any observed errors. The errors appear to be evenly split between manual (23) and autonomous (23), suggesting that one style of flying is not inherently less error inducing than the other. The errors show a slight dip on 9/30/22 and then rise on 10/2/22, possibly consistent with cumulative pilot fatigue reported in [22].

Twenty-seven missions had a total of 46 instances of human error that were correctable by data managers, but delayed or degraded the final data product and thus are considered non-terminal. The details of the non-terminal errors are as follows:

- Data from 27 missions (1, 2, 3, 4, 5, 10, 13,14, 15, 16, 17,18, 19,21, 22, 23, 24, 25, 26, 28, 29, 30 31, 32, 33, 31, 34) was delivered to the data managers in the wrong format, with either a sortie level naming error, top level naming error, or both. This required the data managers to have to manually determine the mission details and manually rename the files before uploading to incident command. While not individually time consuming, the delay was pronounced because the data managers would get the majority of mission data at the same time when squads returned for the evening. The delays in reformatting large batches of data led to delays in supplying incident command with the data in time for their evening planning cycle.
- Ten missions (1, 2, 4, 5, 6, 7, 9, 11, 25, 26) did not have the requested .SRT file from a drone that supported that option. The .SRT file streamlines documenting where the drone has flown.
- Two missions (28,31) were mapping missions flown manually instead of using autonomous software which guaranteed complete coverage of an area and minimized noise. For mission 28, a squad attempted to collect mapping data despite not having been requested or trained to perform mapping. The data from mission 28 was not usable for mapping purposes. For mission 31, a squad was requested to perform and accepted a mapping mission despite not being trained. The collected data was processed into a map of lower quality than expected.
- One mission (19) was an unrequested mission, where the squad self-dispatched to survey an area that was not requested, wasting time and resources.

Four missions had a total of 5 instances of human error that prevented production of the data product and thus are considered terminal. In Missions 8, 14, and 29, the squads encountered a software bug in the autonomous flight and data collection software, which resulted in no usable data; the human error was not re-flying with a different available data collection software package. Mission 25 had two errors that, in combination, resulted in the terminal failure of the mission. That squad neglected to turn on data recording on the drone, then compounded the error by failing to recognize no data was recorded during the post-flight quality assurance step.

D. Human Errors By Flight Style and Activity

As shown in Table III, the 51 errors were associated with either procedure-based (35) or skill-based (16) activities. Procedure-based activities are those related to following expected protocols for the mission (e.g., adhering to the mission tasking), data management (e.g., performing quality assurance, editing data into the required format), or aviation safety (e.g., following checklist). Skill-based activities involve configuring or controlling the drone. Skill-based errors reflect the squad's direct interaction with a specific drone. In contrast, procedure-based errors reflect the relationship between the squad and the larger incident command system which is tasking the drone and consuming the data products. Human-robot interaction has traditionally restricted its focus to squad-drone control. Still, the high number of procedural errors and their negative impact on meeting the mission goals suggests the focus should be broadened beyond only satisfying aviation safety. These errors raise questions as to the impact of training and reminders on enabling squads to temporarily adopt new procedures and on the true level of situation awareness being maintained by pilots during autonomous flight.

The errors in the procedure-based activities were more evenly distributed between autonomous (22) and manual flight styles (13) than skill-based errors, which were 4 and 12, respectively. This is not surprising given that the procedure-based errors appear to cluster around data reporting and quality assurance tasks that occur independently of the actual flight of a drone.

The procedure-based errors may be a side-effect of squads being expected to adhere to new protocols that conflict with their normal ways of working, despite prior training or a deployment within the past 12 months. In terms of data reporting, each agency has their own internal protocols for formatting and delivering data. When working collaboratively for incident command, agencies are expected to temporarily adopt new protocols. In terms of quality assurance, agencies may not explicitly require a quality assurance check of the data before the squad leads the flight location. One reason is that quality assurance is often ignored in manufacturer-provided checklists, as those checklists are oriented towards aviation safety, not mission objectives. Another reason may be that quality assurance is not required for tactical operations where sUAS data is streamed to provide tactical overwatch of an incident or monitor a building fire. However, as with data reporting, quality assurance was emphasized in training for disasters.

These procedural errors appear to be the result of a collision between normal and off-normal practices, but lack of training does not appear to the cause or solution. Agencies participating in multi-agency events such as disasters are explicitly acknowledge in the mutual aid agreement that they temporarily report a different agency and must follow that agency's protocols. At Hurricane Ian, each squad had participated in at least one training exercise or prior deployment within the past 12 months, which explicitly trained them on the data reporting and quality assurance protocols. In addition to this prior experience, the task force leader reviewed the new protocols at the start of the deployment. Furthermore, each mission tasking for a squad included a slip of paper with the expected format of data product to serve as a reminder.

Skill-based errors were notably less frequent than procedure-based errors, and none impacted aviation safety. However, skill-based errors were more severe as they accounted for all (5) terminal errors (errors resulting in mission failure). No procedural-based errors resulted in a terminal error during a mission. The majority (12) of skill-based errors were associated with manual flight, with 4 associated with autonomous flight. The skill-based errors took four forms: did not configure the drone to record the requested secondary .SRT files containing metadata about manual flights (10), did not recognize a software error during autonomous flight (3 terminal), manually flying a mapping mission instead of using autonomous software package (2), and one mission where a squad did not record the primary data during manual flight (1 terminal) and failed to recognize that data loss during post-flight data inspection (1 terminal).

None of the skill-based errors directly impacted aviation safety but their impact on the overall mission success raises concerns for human-robot interaction and training. They suggest that the squads are not deeply familiar with all of the capabilities of their sUAS. This could be an indication of training that is restricted to normative missions. The three

TABLE IV

HUMAN ERROR COUNT BY MISSION PHASE AND FLIGHT STYLE.

			Initiation (13)		Execution (3)		Termination (35)	
Date	# Missions	# Errors	Autonomous	Manual	Autonomous	Manual	Autonomous	Manual
9/29/22	6	10	0	5	0	0	1	4
9/30/22	5	5	1	3	0	0	1	0
10/1/22	14	20	1	1	0	2	11	5
10/2/22	9	16	2	0	0	1	9	4
Total	**34**	**51**	**4**	**9**	**0**	**3**	**22**	**13**

instances where pilots failed to recognize a software error in flight are more troubling because these instances indicate a lack of meaningful supervisory control. In those missions, the sUAS was flying autonomously, but a software bug was incorrectly re-positioning the camera after take-off; this meant that the sUAS flew the correct path to acquire the data and collected images at the correct intervals along the path, but the images were not usable because the camera was not at the correct nadir angle. These were costly terminal errors as the missions were over large areas of intense interest to incident command, and each failure represented a lost day. The pilot can detect the camera tilt by looking at the camera pane in the pilot's real-time display. Failure to do so suggests that pilots are not maintaining true situation awareness during flight; even if they did not use the autonomous software frequently in their normal operations, the camera view was clearly visible on the pilot's display, and the difference between pointing straight down and at the horizon is obvious. The errors also raise fears that the pilots may have over-trust in the sUAS autonomy, assuming that all aspects of the sUAS must be functioning correctly if the navigational autonomy is working.

E. Human Errors by Phase of Mission

The human errors by phase of the mission offer a surprise: The majority of skill-based errors (13) occurred in the Initiation phase, and the majority of procedure-based errors (33) occurred in the Termination phase. The high frequency of skill-based errors in the Initiation phase, particularly as the majority are associated with the manual flight style, suggests that the squads may not be sufficiently familiar with flying for disaster missions despite frequent use in their regular jobs.

As background, a mission activity is divided into three phases: Initiation (e.g., preparation for flight and configuration of any software control settings), Execution (e.g., flight), and Termination (e.g., quality control of the data, file formatting, post-processing, etc.). Since the phases have different functions, but each impacts mission success, it is especially important to document the frequency of errors in the Initiation and Termination phases since prior work has primarily focused on aviation safety in the Execution phase.

Table IV presents the human error count by phase of mission and flight style while Table V shows the error by phase and activity. A comparison of the two tables shows the errors cluster around the activity, not the flight style. Human error in the Initiation phase is solely skill-based (13), with errors in the Termination phase primarily procedure-based

TABLE V

HUMAN ERROR COUNT BY MISSION PHASE AND ACTIVITY.

			Initiation (13)		Execution (3)		Termination (35)	
Date	# Missions	# Errors	Procedure	Skill	Procedure	Skill	Procedure	Skill
9/29/22	6	10	0	5	0	0	5	0
9/30/22	5	5	0	4	0	0	1	0
10/1/22	14	20	0	2	1	1	15	1
10/2/22	9	16	0	2	1	0	12	1
Total	**34**	**51**	**0**	**13**	**2**	**1**	**33**	**2**

(33 out of 35), and the few errors occurring in the Execution phase (3) fairly evenly split between procedure-based (2) and skill-based (1). Furthermore, the skill-based Initiation errors decrease over time while the procedure-based Termination errors increase.

Comparing Table IV and Table V indicates that it was the manual configuration that was the locus of problems in the Initiation phase, not the configuration of the autonomous software despite most squads using that software less frequently. This could indicate that the squads paid more attention to the autonomous software, either because it was new or because the risks of a mishap were perceived to be lower than a more mundane manual flight. It could also indicate that familiarity with manual control led to over-confidence in configuring that flight style.

The errors in the Termination phase occur with both autonomous and manual flight styles, and the magnitude of errors trends with the number of missions each day; essentially, the data from almost every mission was incorrectly handled. Yet, as noted in the Introduction and Sec. III-A, the squads had been trained to use those procedures and were given explicit verbal and written instructions.

IV. DISCUSSION

The preliminary analysis yields four main findings, though it is limited by the nature of using mission logs as the sole source of observing human error.

A. Findings

There was a high incidence of human error in completing missions correctly. Only three missions (8.8%) were completed without errors of any kind, which is troubling, given that errors were likely under-counted. Four of 34 missions (11.7%) ended in terminal failure due to human error, while 31 completed missions (91.2%) evidenced human error. There were almost twice as many procedure-based mistakes (35) as skill-based mistakes (16); the procedure-based mistakes resulted in mission completion delays (i.e., data delivery delays, unavailability of the squad), while a small portion of the skill-based mistakes resulted in terminal failures. Skill-based mistakes were related to configuring autonomous software packages, not direct piloting skills or aircraft safety.

The use of sUAS should be treated as a data-to-decision system, not a drone control problem or an autonomy problem. The high incidence of human error in producing the requested data products indicates the need for research, development, and training in small UAS to consider the entire mission and mission objectives, not just the aviation safety or autonomous navigation components. As such, designs and training should include the Initiation and Termination phases, not just the Execution flight phase. The error counts indicate that adding an intelligent assistant for configuring for specific missions, reminding the operator of mission constraints not typically included on a manufacturer's checklist, and providing interactive troubleshooting may offer more practical value than increasing platform flight autonomy.

TABLE VI

HUMAN ERROR COUNTS BY INSTITUTION.

Date	# Missions	# Errors	Insurance/ Academic (2 squads)	Fire Rescue (3 squads)	Law Enforcement (5 squads)
9/29/22	6	10	2	8	0
9/30/22	5	5	3	2	0
10/1/22	14	20	1	3	16
10/2/22	9	16	1	9	6
Total	**34**	**51**	**7**	**22**	**22**

Training for off-normal, inter-agency deployments may be need to be expanded or conducted more regularly. Table VI shows that the eight squads from fire rescue and law enforcement agencies had more instances of human error than the university or insurance team. The fire rescue and law enforcement agencies presumably use their sUAS far more frequently than the non-governmental squads, though that makes them predisposed to revert to their own procedures. Regardless of the cause, the high error count suggests that training should focus on preparing or refreshing a broad set of skills beyond the normative routine. The training, and possibly the organizational culture, should help the squad become comfortable with new procedures and able to adapt to the requests and practices of whatever agency has jurisdiction for the disaster.

The root cause of the procedure-based human errors merits further investigation. The mission logs do not provide insight into the underlying causes of the procedure-based errors. It is intriguing that while the count of human error for each mission per day remained relatively constant (1.7, 1.0, 1.4, 1.7 errors/mission for each of the four days), there was a notable shift in errors by activity. Over time, the number of procedure-based errors increased. If the cause was lack of familiarity, poor training, or other issues associated with operating robots in an off-normal situation [23], the squads would be expected to have acclimated to novel processes. Instead it appears they performed worse over time. Another explanation may be fatigue.[22] It would be interesting to examine the number of missions and pattern of errors over time by squad and activity to determine if fatigue might be an explanation and if procedure-based errors are more likely than skill-based errors.

B. Limitations

Using mission logs for a post hoc analysis is not ideal for quantifying human error and determining root causes. However, it provides a lower bound on obvious errors made during a mission, giving some insight into what is happening during missions. As shown in this paper, even coarse error data can lead to discoveries: that procedural errors are as important as skill-based errors. It also avoids disrupting the life-saving response with more traditional human-robot interaction research methods of data capture (e.g., adding HRI observers, conducting interviews or surveys, and capturing biometrics).[21]

Examining mission logs for indications of human errors offers three advantages. One is that it provides real-world data on human error during sUAS deployments without inserting observers or new procedures (e.g., interviews) into operations, both of which pose both logistical and ethical issues as described in [24]. Second, examination could be done in real-time, where the task force management can see errors and attempt to correct them within one operational shift through mechanisms such as just-in-time training to squads with skill deficits, reinforcing procedures during morning briefings and evening after-action reports, etc. Third, the detection of squads exhibiting errors or an increasing frequency of errors could be automated and depersonalized. This would enable the task force leader to determine the source of errors and mitigate appropriately (e.g., that the squad was showing signs of fatigue and needs to be rotated out) and would diffuse the perception that squads were being selectively criticized thus undermining team cohesion.

The disadvantage is that the mission logs are a historical management record, not an intentional capture of human error during the mission. As a result, there are likely to be gaps and data that lead to incorrect inferences. The logs do not capture the occurrence of pilot errors during the flight itself, so there could be a high count of errors during the Execution phase. The mission logs could not capture any errors that did not directly impact delivery of the desired data products to the incident command. As a result, mission logs are not helpful in ascertaining the root causes of human error, though they can alert the presence of errors and the need for further investigation.

V. Conclusions

This paper reports on human error in the use of small UAS during the Hurricane Ian response, where 31 of the 34 missions flown had at least one error discernible from the mission logs. In addition to quantifying the errors, the paper makes three contributions to the field of disaster robotics described below. From a practitioner's viewpoint, the paper shows the need for inter-agency training on off-normal flying styles and procedures.

It reinforces the premise that disaster robotics is more than the direct operation of a robot, in this case, a drone. The robot is being used as a tool to acquire data essential to the disaster management enterprise; thus, the robot is an element within a data-to-decision process. The paper highlights that current robot systems do not facilitate data-to-decision work processes, as they tend to be oriented towards aviation safety and flight autonomy (i.e., the Execution phase), not configuration (Initiation phase), or automating quality assurance or simplifying mission-specific data packaging (Termination phase). By not explicitly acknowledging that humans are involved in the process, roboticists may overlook opportunities to use AI to assist error-prone operators, especially when working in off-normal situations such as a disaster.

The paper merged two existing human-robot failure taxonomies to produce a more comprehensive model for field robotics. The new partitioning of human error also includes the activity, either procedure- or skill-based, associated with the failure. The analysis in this paper illustrated how the refinement of errors by activity was of value.

The paper also offers a new methodology for unobtrusively capturing human error by examining mission logs. Although this method does not necessarily provide explanations or root causes, it is quantitative, and the measures could be monitored in real-time, enabling timely mitigations. The mission logs, with squads and locations anonymized, are available upon request.

Acknowledgment

The authors thank David Merrick, Task Force Leader for FL-UAS-1, and the Florida State University Center for Disaster Risk Policy Disaster Incident Response Team for their assistance.

References

[1] T. Manzini, R. Murphy, and D. Merrick, "Quantitative data analysis: Crasar small unmanned aerial systems at hurricane ian," *2023 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2023.

[2] O. Fernandes, R. Murphy, D. Merrick, J. Adams, L. Hart, and J. Broder, "Quantitative data analysis: Small unmanned aerial systems at hurricane michael," in *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Conference Proceedings, pp. 116–117.

[3] O. Fernandes, R. Murphy, J. Adams, and D. Merrick, "Quantitative data analysis: Crasar small unmanned aerial systems at hurricane harvey," in *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Conference Proceedings, pp. 1–6.

[4] Q. Xiao, Y. Li, F. Luo, and H. Liu, "Analysis and assessment of risks to public safety from unmanned aerial vehicles using fault tree analysis and bayesian network," *Technology in Society*, vol. 73, p. 102229, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160791X23000349

[5] M. Ghasri and M. Maghrebi, "Factors affecting unmanned aerial vehicles' safety: A post-occurrence exploratory data analysis of drones' accidents and incidents in australia," *Safety Science*, vol. 139, p. 105273, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753521001181

[6] H. E. Özyörük, "Systematic analysis and classification of the literature regarding the impact of human factors on unmanned aerial vehicles (uav)," *Journal of Aviation*, vol. 4, no. 2, pp. 71–81, 2020.

[7] H. Fesenko and V. Kharchenko, "Determining the optimum number of single operator controlled unmanned aerial vehicles for npp monitoring missions: Human error issues," in *2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology*, 2018, Conference Proceedings, pp. 711–714.

[8] T. Shmelova, Y. Sikirda, and Y. Kovalyov, "Decision making by remotely piloted aircraft system's operator," in *2017 IEEE 4th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, 2017, pp. 92–99.

[9] A. Hobbs, *Chapter 16 - Unmanned Aircraft Systems*. San Diego: Academic Press, 2010, pp. 505–531.

[10] J. Carlson and R. Murphy, "How ugvs physically fail in the field," *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 423–437, 2005.

[11] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in Psychology*, vol. 9, 2018.

[12] S. A. Shappell and D. A. Wiegmann, "Applying reason: The human factors analysis and classification system (hfacs)," *Human Factors and Aerospace Safety*, vol. 1, no. 1, pp. 59–86, 2001.

[13] D. A. LaMonica, K. Drnec, and M. E. Miller, "Employing mbse to assess and evaluate human teaming in military aviation command and control," in *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, 2022, Conference Proceedings, pp. 1–7.

[14] G. Steinbauer, "A survey about faults of robots used in robocup," in *RoboCup 2012: Robot Soccer World Cup XVI 16*. Springer, 2012, Conference Proceedings, pp. 344–355.

[15] K. H. Seo and J. H. Lee, "The emergence of service robots at restaurants: Integrating trust, perceived risk, and satisfaction," *Sustainability*, vol. 13, no. 8, p. 4431, 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/8/4431

[16] L. Tian and S. Oviatt, "A taxonomy of social errors in human-robot interaction," *J. Hum.-Robot Interact.*, vol. 10, no. 2, p. Article 13, 2021. [Online]. Available: https://doi.org/10.1145/3439720

[17] X. Leo and Y. E. Huh, "Who gets the blame for service failures? attribution of responsibility toward robot versus human service providers and service firms," *Computers in Human Behavior*, vol. 113, p. 106520, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563220302727

[18] S. Tolmeijer, A. Weiss, M. Hanheide, F. Lindner, T. M. Powers, C. Dixon, and M. L. Tielman, "Taxonomy of trust-relevant failures and mitigation strategies," in *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2020, Conference Proceedings, pp. 3–12.

[19] A. Sutcliffe and G. Rugg, "A taxonomy of error types for failure analysis and risk assessment," *International Journal of Human-Computer Interaction*, vol. 10, no. 4, pp. 381–405, 1998.

[20] R. R. Murphy, "Mission-oriented ontology for small unmanned aerial systems for disasters," in *International Workshop on Ontologies and Standards for Robotics and Automation (WOSRA 2024 @ICRA2024)*, Conference Proceedings.

[21] S. C. Peres, R. Mehta, and R. Murphy, "Conducting human-robot interaction research during three real-world disasters," in *HRI 2021 Workshop on Metrics and Test Methods*, Conference Proceedings.

[22] S. C. Peres, R. K. Mehta, and R. R. Murphy, "Water, lava, and wind," *Interaction Studies*, vol. 24, no. 3, pp. 335–361, 2023. [Online]. Available: https://www.jbe-platform.com/content/journals/10.1075/is.22048.per

[23] R. R. Murphy, "Humans and robots in off-normal applications and emergencies," in *10th International Conference on Applied Human Factors and Ergonomics (AHFE 2019)*, Conference Proceedings.

[24] ——, "Participant observer ethnography in disaster robotics," in *Ethnography for HRI: Embodied, Embedded, Messy and Everyday @HRI 2024*, Conference Proceedings.