# AN EVALUATION OF UAS PILOT WORKLOAD AND ACCEPTABILITY RATINGS WITH FOUR SIMULATED RADAR DECLARATION RANGES

Jillian N. Keeler[1], R. Conrad Rorie[1], Kevin J. Monk[1], Garrett G. Sadler[1,2], Casey Smith[1]
NASA Ames Research Center[1], San José State University[2]

Currently, minimum operational performance standards (MOPS) are being developed for a broader range of unmanned aircraft system (UAS) platforms, including smaller UAS that will feature onboard sensors that are low in size, weight, and power, otherwise known as low SWaP. The low SWaP sensors used to detect non-cooperative traffic will have limited declaration ranges compared to those designed for medium-to-large UAS. A human-in-the-loop (HITL) study was conducted examining four possible radar declaration ranges (i.e., 1.5 NM, 2 NM, 2.5 NM, and 3 NM) for a potential low SWaP sensor with a detect and avoid (DAA) system encountering various non-cooperative encounters in Oakland Center airspace. Participants had lower workload, particularly workload associated with temporal demand and effort, in scenarios that featured larger detection ranges. Furthermore, participants reported better ability to remain DAA well clear within the larger declaration range conditions, specifically with the 2.5 NM and 3 NM conditions.



## INTRODUCTION

The integration of unmanned aircraft systems (UAS) into the national airspace system (NAS) has been a massive, multi-agency effort for more than a decade. The research and development over that time period has focused primarily on defining operational requirements for medium-to-large UAS flying through Class D, E, or G airspace (RTCA, 2017). One major emphasis of the effort has been defining "detect and avoid" (DAA) standards that can interoperate with the reciprocal standard of the "see and avoid" requirement in traditional aviation. Since a UAS pilot cannot visually identify aircraft that the UAS may be in conflict with, onboard sensors such as Automatic Dependent Surveillance-Broadcast (ADS-B) and an air-to-air radar (ATAR) are used to detect intruders. ADS-B is able to detect "cooperative" aircraft (i.e., aircraft with an operational transponder), whereas an ATAR is used to detect "non-cooperative" aircraft (i.e., aircraft without an operational transponder). A DAA system can take information from these sensors and provide alerting and maneuver guidance to UAS pilots based upon a set of defined DAA well clear volumes. These volumes can vary based on intruder equipage (i.e., cooperative or non-cooperative) and operational area (e.g., en-route, terminal area).

Previous studies have looked at DAA system requirements for larger UAS operating in en-route and terminal environments (Fern et al., 2018; Rorie et al., 2017; Rorie et al., 2019). More recently, DAA research has focused on developing Minimum Operational Performance Standards (MOPS) for a broader range of UAS types, including smaller UAS that will need onboard sensors classified as low in size, weight, and power (SWaP) (RTCA, 2019). These aircraft will not be able to support the type of sensors that were developed for the larger UAS (e.g., RQ-4 Global Hawk). As a result, UAS equipped with low SWaP sensors will have smaller surveillance ranges than previously assumed. It is therefore critical that low SWaP-specific DAA system requirements (e.g., well clear definition, alerting requirements) be developed and tested to support the pilot's ability to maintain DAA well clear (DWC) of these intruders with a more limited declaration range.

A previous human-in-the-loop (HITL) study examined two different non-cooperative DWC definitions for a low SWaP-equipped UAS (Monk et al., 2020a). The low SWaP radar declaration range (RDR) simulated in this study was fixed at 3.5 nautical miles (NM) with a limited field of regard. The goal of the simulation was to determine whether either of these definitions out-performed the other (i.e., improved the pilot's ability to maintain DWC). The results showed nearly equivalent performance. One of the definitions, however, maximized the amount of alerting time available to the pilot, which could help facilitate higher rates of air traffic control (ATC) coordination. This finding was consistent with a fast-time simulation that found that a DWC with a horizontal miss distance of 2,200 ft and a vertical miss distance of 450 ft was suitable for low SWaP-equipped UAS (Wu et al., 2018). Both of these studies suggested further work was needed to investigate smaller declaration ranges than 3.5 NM with the newly identified low SWaP DWC definition.

The following HITL study examined declaration ranges shorter than 3.5 NM to help identify a minimally-acceptable declaration range for a low SWaP sensor. The objective results of this HITL are reported by Monk et al. (2020a). The current paper focuses on the subjective feedback received from pilots over the course of this study.

## METHOD

### Experimental Design

The experimental design consisted of low SWaP radar declaration range as the main within-subject variable and ownship speed as a between-subjects variable. Low SWaP RDR was counterbalanced across scenarios and could be either 1.5 NM, 2 NM, 2.5 NM, or 3 NM. Ownship speed was split amongst participants and could either be slow (i.e., 60 knots [KTS]) or fast (i.e., 100 KTS). Conflict type varied within each trial by either the equipage (i.e., one cooperative and four non-cooperative) and the closure rate of the intruder. Closure rate was manipulated by changing the intruder speed (i.e., 100 versus 170 KTS) and the approach angle relative to ownship (i.e., head-on or 90° crossing).

## Participants

Nine participants ($M$ = 35 years old) were recruited from a pool of active-duty UAS pilots. Reported military non-combat hours averaged at 356 hours, whereas military combat had an average of 822 hours. Ownship speed was randomly assigned to participants. Five participants experienced the slow ownship condition and four participants experienced the fast ownship condition.

## Apparatus

*Ground Control Station (GCS)*. The GCS used for this simulation was Vigilant Spirit Control Station (VSCS), which was developed by the Air Force Research Laboratory (AFRL) and consisted of two displays (Feitshans et al., 2008). The main center display, known as the Tactical Situation Display (TSD), showed an airspace map overlay, aircraft controls, and aircraft control state. The secondary display, known as the tote board, displayed additional aircraft information (e.g., relative fuel burn per hour), contingency event checklists (e.g., generator failure), and a mission command chat room. Vigilant Spirit Simulation (VS Sim) acted as an event generation tool for the planned intruder encounters and was utilized by researchers in a room separate from the GCS. Voice communication with a researcher acting as an air traffic controller for Oakland Center was conducted via push-to-talk headsets and a voice IP server.

*Simulated Aircraft and Airspace.* Participants flew a generic RQ-7 Shadow at a mission altitude of 8,000 ft MSL and a cruise speed of either 60 or 100 KTAS along a mission route following a racetrack pattern. The simulated UAS was equipped with ADS-B, which detected cooperative aircraft within 20 NM of the ownship, and a low SWaP radar, which detected non-cooperative aircraft at either 1.5 NM, 2 NM, 2.5 NM, or 3 NM depending on the condition. The low SWaP radar also had a restricted field of regard, with an azimuth of ±110° off the nose of the aircraft and an elevation of ± 15°.

The mission was conducted within Oakland Center Class E airspace, with background traffic injected through VS Sim. An active temporary flight restriction (TFR) zone was placed south of the route to simulate similar procedures that would be used for a later live flight test using the same airspace overlay and conditions.

*DAA System.* The DAA system used for the simulation incorporated NASA's Detect and Avoid Alerting Logic for Unmanned Systems (DAIDALUS) and Java Architecture for DAA Extensibility and Modeling (JADEM) (Muñoz et al., 2015). For the iconography and associated aural alerts for each DAA alert level, see Table 1.

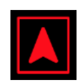
At the DAA Warning Alert level, pilots needed to make an immediate action to maintain DWC (e.g., vertical maneuver, horizontal maneuver, or a combination of both). ATC coordination prior to the maneuver was not required, as pilots only had up to 30 seconds to avoid a loss of DAA well clear (LoDWC). Pilots were encouraged to advise ATC of their maneuver following successful avoidance of the intruder.

At the Corrective DAA Alert level, a corrective action was required to maintain DWC but required coordination with ATC prior to executing the avoidance maneuver. In this scenario, pilots had up to 60 seconds to avoid a LoDWC.

A Preventive DAA Alert required no action to maintain DWC and, instead, acted as a notice to monitor an aircraft for a potential escalation to a higher DAA alert. If this alert was active, the ownship and other aircraft were separated by at least 500 ft vertically with a max separation of 700 ft.

Guidance traffic was indicated by a solid white intruder symbol with no accompanying aural alert whenever an aircraft was generating DAA guidance bands outside of the ownship's current course. Basic traffic was all traffic that was detectable but did not satisfy any of the other intruder types. It was indicated by a hollow intruder icon and featured no aural alert.

Table 1. Detect and Avoid (DAA) Alerting Structure

| Symbol | Name | Time to LoDWC | Aural Alert Verbiage |
|---|---|---|---|
|  | DAA Warning Alert | 30 s | "Traffic, maneuver now. Traffic, maneuver now." |
|  | DAA Corrective Alert | 60 s | "Traffic, avoid." |
|  | DAA Preventative Alert | N/A | "Traffic, monitor." |
|  | Guidance Traffic | N/A | N/A |
|  | Basic Traffic | N/A | N/A |

## Procedures

*Training.* At the beginning of the day, pilots received a general briefing on the project and VSCS. Following the briefing, a warm-up session lasting roughly 30 minutes on station was completed to familiarize the pilot with controls and initiating maneuvers. Once familiarized with VSCS, pilots received a briefing on the DAA system and associated guidance. Preceding each experimental trial, pilots received a 20-minute training session with DAA encounter examples and whichever RDR condition they were about to experience.

*Experimental Trials*. Participants completed four separate experimental trials over the course of one day, simulating each of the four RDRs. Trials lasted at least 45 minutes. Five scripted encounters set to lose DWC were injected throughout the course of each trial. Encounters were counterbalanced across trials to prevent possible learning effects. Pilots were to respond accordingly to intruders based upon the level of DAA alerting and guidance being presented. Following maneuver avoidance, pilots were to contact ATC to receive approval to return to course. Additionally, pilots responded to secondary tasks such as scripted situational awareness chat messages (e.g. “What is your current bearing and range to the next waypoint on route?”) and failure events (i.e., header tank overpressure or generator failure). Responses to failure events included completing the associated checklist.

Following the completion of a trial, participants completed a questionnaire which included the NASA task load index (TLX) and ratings of statements meant to assess the acceptability of a given RDR. For each statement, participants used a rating scale where ‘1’ indicated strong disagreement, ‘3’ represented a neutral point, and ‘5’ captured strong agreement. Following completion of all four RDR conditions, participants completed a post-simulation questionnaire which gauged their overall experience with the DAA system. A verbal debrief was also completed with participants to capture any additional feedback.

## RESULTS

To determine if any components of the NASA TLX scores or post-trial questionnaire responses differed as a result of the simulation’s manipulations, separate 2 (Ownship Speed: Slow or Fast) x 4 (RDR: 1.5 NM, 2 NM, 2.5 NM, 3 NM) mixed analysis of variance (ANOVAs) were run. All pairwise testing used individual paired samples t-tests for the within-subjects variable of RDR. Holms-Bonferroni corrections were used to test significance of all six possible pairwise comparisons.

### NASA TLX

None of the NASA TLX components featured significant main effects of ownship speed or significant interactions between RDR and ownship speed, *p*s > .05. Additionally, the mental demand and physical demand dimensions did not vary significantly as a result of RDR, *p*s > .05. Components such as performance, $F(1.66, 11.60)$, = 5.85, $p = .02$, and frustration, $F(3, 21) = 6.07$, $p = .004$, featured significant main effects of RDR, but failed to produce any significant pairwise comparisons when utilizing corrected p-values. The following results from the NASA TLX components discuss all significant main effects of RDR that featured at least one significant pairwise comparison.

Participants reported varying levels of temporal demand across most of the RDR conditions, $F(3, 21) = 11.41$, $p < .001$. The 1.5 NM condition had higher temporal demand scores than the 2.5 NM, $t(8) = 3.90$, $p = .005$, and 3 NM conditions, $t(8) = 3.80$, $p = .005$. However, the 1.5 NM condition did not differ in temporal demand scores compared to the 2 NM condition, $t(8) = 1.51$, $p > .05$. Additionally, the 2.5 NM and 3 NM conditions had similar scores, $t(8) = .43$, $p > .05$.

For the main effect of RDR on the effort component scores, $F(3, 21) = 4.19$, $p = .02$, pairwise testing revealed significant differences in effort between the 1.5 NM and 2.5 NM RDRs, such that the 1.5 NM condition had higher effort scores than those found for the 2.5 NM, $t(8) = 3.59$, $p = .007$. However, the 1.5 NM and 3 NM conditions failed to show significant differences when corrections for pairwise testing were applied, $t(8) = 2.73$, $p = .03$. Additionally, the 1.5 and 2 NM conditions did not significantly differ in perceived effort expended, $t(8) = 1$, $p > .05$. Similarly, the 2 NM condition had almost equal scores compared to both the 2.5, $t(8) = 2$, $p > .05$, and 3 NM conditions, $t(8) = 2.12$, $p > .05$. Finally, the 2.5 and 3 NM conditions also did not significantly differ from each other, $t(8) = .82$, $p > .05$.

When examining NASA TLX composite scores, $F(3, 21) = 9.76$, $p < .001$, higher NASA TLX scores were observed in the 1.5 NM condition compared to the 2 NM, $t(8) = 2.30$ $p = .05$, the 2.5 NM, $t(8) = 4.12$, $p = .003$, and the 3 NM conditions, $t(8) = 3.55$, $p = .008$. Nonetheless, comparisons between the 2 NM condition and both the 2.5 NM, $t(8) = 2.92$, $p = .02$, and the 3 NM conditions, $t(8) = 2.80$, $p = .02$, failed to reach significance after pairwise comparison corrections were applied. No significant difference in NASA TLX score was observed between the 2 NM, 2.5 NM, and 3 NM conditions, $t(8) = 1.15$, $p > .05$ (see Table 2).

Table 2. Average NASA TLX Scores by Radar Declaration Range (RDR)

| | **Radar Declaration Range (RDR)** | | | |
|---|---|---|---|---|
| **TLX Component** | **1.5 NM** *M* (*SD*) | **2 NM** *M* (*SD*) | **2.5 NM** *M* (*SD*) | **3 NM** *M* (*SD*) |
| **Mental** | 3.33 (1.73) | 3 (1.50) | 2.67 (1.23) | 2.33 (1) |
| **Physical** | 1.89 (.93) | 1.78 (.97) | 1.67 (.71) | 1.67 (.87) |
| **Temporal** | 4.78† (1.17) | 4.33 (1.23) | 2.89† (.78) | 2.78† (.67) |
| **Performance*** | 3.22 (1.79) | 2.22 (1.09) | 2 (1) | 1.78 (.67) |
| **Effort** | 3.78† (1.48) | 3.33 (1.58) | 2.67† (1.12) | 2.33 (.71) |
| **Frustration** | 3.67 (2) | 2.78 (1.72) | 2.22 (1.30) | 2 (1) |
| **Composite** | 20.67† (6.98) | 17.44† (6.17) | 14.11† (3.82) | 12.89† (2.71) |

**Performance ranges from perfect (‘1’) to failure (‘7’)*
*† Indicates significant pairwise comparisons*

**Post-Trial Questionnaire**

Once again, no significant main effects of ownship speed or significant interactions between ownship speed and RDR were observed, *p*s > .05. The following results discuss the significant main effects of RDR that were found with post-trial questions that featured at least one significant pairwise comparison.

RDR impacted whether participants felt that the corrective DAA alerts provided enough time to coordinate with ATC, $F(3, 21) = 6.20$, $p = .003$. Participants reported better time for coordination at the corrective level for the 3 NM condition as compared to the 2 NM condition, $t(8) = -4.91$, $p = .001$. No significant differences were found between the 1.5 NM condition and the 2 NM, $t(8) = 1$, $p > .05$, or the 2.5 NM conditions, $t(8) = -1.51$, $p > .05$. This was also true for the 2.5 NM and 3 NM conditions, $t(8) = -1.84$, $p > .05$.

Perception of ability to maintain DAA well clear varied with RDR, $F(3, 21) = 9.59$, $p < .001$, such that the 1.5 NM condition was viewed as producing significantly less ability to maintain DAA well clear than the 2.5 NM, $t(8) = -4.26$, $p = .003$ and 3 NM conditions, $t(8) = -5.29$, $p = .001$. Conversely, differences between the 2 NM condition and both the 1.5 NM, $t(8) = -2.87$, $p = .02$, and the 3 NM, $t(8) = -2.83$, $p = .02$, in ability to remain well clear failed to be significant after pairwise comparison corrections. The 2 NM and 3 NM condition were seen as possessing similar ability to maintain well clear, $t(8) = -.69$, $p > .05$, as well as the 2.5 NM and 3 NM conditions, $t(8) = -1.84$, $p > .05$.

When asked whether they would feel comfortable flying with a DAA system with a given RDR in the NAS, $F(3, 21) = 8.51$, $p = .001$, both the 2.5 NM, $t(8) = -3.78$, $p = .005$, and 3 NM RDRs, $t(8) = -4.13$, $p = .003$, were seen as being more comfortable than the 1.5 NM condition. However, ratings for the 2 NM condition compared to both the 2.5 NM, $t(8) = -2.86$, $p = .02$, and the 3 NM RDRs, $t(8) = -2.68$, $p = .03$, failed to pass significance testing after pairwise comparison corrections. Meanwhile, the 2.5 NM and 3 NM RDR conditions had relatively similar ratings, $t(8) = .43$, $p > .05$ (see Table 3).

Table 3. Average Post-Trial Responses by Radar Declaration Range (RDR) (1 = Strong Disagreement, 3 = Neutral, 5 = Strong Agreement)

| | **Radar Declaration Range (RDR)** | | | |
|---|---|---|---|---|
| **Post-Trial Questions** | **1.5 NM** *M* (*SD*) | **2 NM** *M* (*SD*) | **2.5 NM** *M* (*SD*) | **3 NM** *M* (*SD*) |
| **Coordination with ATC** | 3.67 (1.23) | 3.11† (1.05) | 4.11 (.93) | 4.56† (.53) |
| **Ability to Maintain DWC** | 2.55† (1.01) | 3.33 (1.12) | 3.56† (.73) | 4† (.71) |
| **Comfort in NAS** | 2.67† (1.58) | 3.22 (1.30) | 4.33† (.50) | 4.22† (.97) |

† *Indicates significant pairwise comparisons*

**Post-Simulation Questionnaire and Debrief**

Following the completion of all four trials, participants completed a post-simulation questionnaire. When asked which RDR they experienced during the simulation could be considered minimally acceptable, responses were somewhat split. Five participants indicated that the 2.5 NM RDR would be minimally acceptable, whereas four participants selected the 2 NM RDR. The four participants who experienced fast ownship speeds all selected the 2.5 NM RDR, with one additional participant who experienced slow ownship speed. All other participants who experienced slow ownship speed chose the 2 NM RDR.

Additionally, participants provided clarifying comments regarding their choices during the verbal debrief session. One participant noted that the 1.5 NM RDR “does not allow enough time for detection, orientation, coordination and maneuvering.” Another commented on the aircraft flown for the simulation, noting that it is “not a high-performance machine” and that “warning times and buffers need to reflect the aircraft’s real time capabilities to actually avoid collisions.” Additionally, some participants noted that smaller RDRs could be potentially supported by automated features (i.e., auto-fill of maneuver within the aircraft control panel once DAA alerting is triggered that would require the pilot to manually upload), thus reducing the time added by having the pilot interpret the DAA guidance and key in the maneuver.

## DISCUSSION

The findings of our study examining various radar declaration ranges indicated lower acceptability of smaller declaration ranges for a potential low SWaP radar. Participants reported higher workload in the lower declaration ranges, including significant demands on temporal resources, general effort, and overall workload in the 1.5 NM condition. Generally, workload remained lower for both the 2.5 NM and 3 NM conditions. Smaller declaration ranges also provided less adequate time to respond to a LoDWC and limited coordination time with ATC. Participants reported that they would feel safer flying in the current NAS with declaration ranges of 2.5 NM or 3 NM. However, ownship speed had an impact on which RDR participants found to be minimally acceptable, with the vast majority of those who experienced slower ownship speeds selecting the 2 NM RDR and participants who experienced faster ownship speeds favoring the 2.5 NM RDR. This divide in preference is most likely due to the fact that those who had slow ownship speeds never experienced the worst-case scenario (i.e., fast ownship speed encountering a fast head-on intruder).

However, the presented results are constrained by a series of factors. Only the participant and ATC communicated on the common frequency, unlike larger-scale HITL simulations that utilize pseudopilots to provide realistic background radio chatter typical of the Oakland Center radio frequency (Monk et al., 2020a). This led to shortening of both the time it took for a participant to request a maneuver to avoid a LoDWC and receive clearance for that maneuver from the confederate ATC. Furthermore, the present data must be understood alongside data of pilots' objective performance with these declaration ranges. While the authors feel that subjective data offers critical insight into acceptability of shorter sensor ranges, ultimately objective performance is needed to provide insight into how safely pilots can perform the DAA function at these levels. The objective results from this study will be published separately (Monk et al., 2020b), as will the results of a flight test that was performed recently and emulated a low SWaP radar with a 2.5 NM RDR (Vincent et al., 2020).



## REFERENCES


Feitshans, G. L., Rowe, A. J., Davis, J. E., Holland, M., & Berger L. (2008). Vigilant spirit control station (VSCS) – 'The face of COUNTER'. In *AIAA Guidance, Navigation and Control Conference Exhibition.* DOI: 10.2514/6.2008-6309

Fern, L., Rorie, R. C., Roberts, Z., & Monk, K. (2018). An exploratory evaluation of UAS detect and avoid operations in the terminal environment. In *2018 Aviation Technology, Integration, and Operations Conference* (p. 2874). DOI: 10.2514/6.2018-2874

Monk, K., J., Rorie, R. C., Keeler, J. N., & Sadler, G. G. (2020a). An examination of two non-cooperative detect and avoid well clear definitions. In *AIAA Aviation 2020 Forum and Exposition*.

Monk, K. J., Rorie, R. C., Sadler, G., Keeler, J., & Smith, C. (2020b). UAS pilot performance comparisons with different low size, weight, and power sensor declaration ranges. In *39th Digital Avionics Systems Conference*.

Muñoz, C., Narkawicz, A., Hagen, G., Upchurch, J., Dutle, A., Consiglio, M., & Chamberlain, J. (2015). DAIDALUS: Detect and avoid alerting logic for unmanned systems. In *34th Digital Avionics Systems Conference* (p. 55A1-1). DOI: 10.1109/DASC.2015.7311421

Rorie, R. C., Fern, L., Monk, K., Santiago, C., Shively, R. J., & Roberts, Z. S. (2017). Validation of minimum display requirements for UAS detect and avoid system. In *17th AIAA Aviation Technology, Integration, and Operations Conference* (p. 4483). DOI: 10.2514/6.2017-4483

Rorie, R. C., Monk, K. J., Brandt, S. L., Roberts, Z. S., & Fern, L. (2019). A comparison of two terminal area detect and avoid well clear definitions. In *AIAA Aviation 2019 Forum* (p. 3182). DOI: 10.2514/6.2019-3182

RTCA (2017). DO-365: Minimum Operational Standards for Unmanned Aircraft Systems (UAS) Detect and Avoid Systems (DAA), Washington, D.C.: RTCA.

RTCA (2019). Paper No. 225-19/PMC-1941: Minimum Performance Standards for Unmanned Aircraft (Rev 8), Washington, D.C.: RTCA.

Vincent, M., Rorie, R. C., Monk, K. J., Sadler, G., Keeler, J., & Smith, C. (2020). UAS integration into the NAS flight test 6: Full mission. In *39th Digital Avionics Systems Conference*.

Wu, M. G., Cone, A. C., Lee, S., Chen, C., Edwards, M. W., & Jack, D. P. (2018). Well clear trade study for unmanned aircraft detect and avoid with non-cooperative aircraft. In *AIAA Aviation 2018 Forum* (p. 2876). DOI: 10.2514/6.2018-2876