Qualitative and Quantitative Risk Assessment of Urban Airspace Operations

Leonid Sedov
Communications and Transport Systems, ITN
Linköping University
Norrköping, Sweden

Valentin Polishchuk
Thibault Maury
ENAC
Toulouse, France

Maria Ulloa
Universidad Politécnica de Madrid
Madrid, Spain

Darya Lykova
Institut Le Rosey
Rolle, Switzerland

Abstract—Specific Operations Risk Assessment (SORA) is a qualitative methodology for assessing risks of drone operations. In this paper, SORA is compared to and complemented with quantitative estimations of the risk (earlier called HFRM: High-fidelity risk modeling). We highlight intrinsic shortcomings of both SORA and HFRM, and show how HFRM may help to deal with SORA’s ambiguities. We do not have a recipe to remedy HFRM’s drawbacks with the help of SORA, but suggest a possible regulatory fix to HFRM, addressing its deficiency.) With its focus on ground risk, this paper complements the works of TU Dresden which suggested integrating “agent simulation as air risk assessment in SORA” [Fricke et al., ATM Seminar 2021] and of SESAR’s ER4 BUBBLES project “proposing a quantitative risk analysis which enhances or replaces the qualitative model of SORA” (also for the air risk) [BUBBLES Deliverable 4.1]; we also connect to CORUS observations on SORA shortcomings and the use U-space services for addressing them. Our work advocates for stricter regulations, including digitalization and automation not only in definitions, but also in mandates/requirements. Our arguments are illustrated on simple synthetic cases and on real-world experimental examples from urban areas.

Keywords—Unmanned Aerial Systems; High-fidelity risk modeling; Specific Operations Risk Assessment; Ground risk; Air risk

I. INTRODUCTION

Risk assessment is at the core of the approval process for drone operations. Current risk assessment methods may be split into two broad categories:

- **Qualitative.** Specific operations risk assessment (SORA) [1] pages 11-32] starts from initial ground and air risk classes (GRC and ARC) and reduces the risks by considering various mitigations.

- **Quantitative.** High-fidelity risk modeling (HFRM) [2][12] estimates the expected fatality rate (EFR) of the operation (the EFR should not exceed 1 fatality per one million flight hours).

SORA allows one to perform risk assessment essentially without using a computer because no coding or data processing is required. To quote [3], “SORA steps … provide a risk assessment that does not require substantial knowledge of how to determine risks and necessary measures to mitigate them”. The methodology (sometimes extended to holistic MEDUSA approach [13]) is spreading quickly in Europe as an industry standard to a wide circle of drone operators: the number of performed SORAs is thousands and counting.

At the same time, quantitative assessments of drone operations are scarce: to our knowledge, the only comparison between SORA and HFRM on two scenarios was done in [3]. For both scenarios the output of SORA and HFRM was the same – operation approved, the risk is sufficiently low. In this paper we show how SORA may benefit from HFRM and computational HFRM-based solutions for finding Pareto-optimal routes that tradeoff efficiency and risk in a precise, quantifiable way. We also advocate for Pareto optimality as a barrier to a potential misuse of low EFR as the sole criteria for operation approval.

The rest of the text is organized as follows. Sections [II] and [III] describe the quantitative methodologies which we use for air and ground risk assessment respectively; we also survey related prior research, putting our work in the context. Sections [IV] and [V] describe the drawbacks of and suggested remedies for qualitative and quantitative risk assessment respectively. We conclude in Section [VI].

II. AIR RISK

Ideas for quantification of *air risk* in SORA were most recently expressed in Deliverable 4.1 [14] of SESAR’s ER4 project BUBBLES (saying “The SORA methodology will be guideline but the goal is to improve the qualitative methodology of SORA with a quantitative methodology, since that will allow to estimate a value for the RWC [Remain Well Clear] separation.”) and ATM Seminar paper [11] (saying about their simulations that “TUD [Dresden University of Technology] suggest integrating this type of agent simulation as air risk assessment in SORA”). In this paper we elucidate similar ideas for *ground risk*; see the next section. Meanwhile, in this section we briefly review existing approaches to quantification of air risk and present calculations of EFR due to conflicts with ADSB-equipped (Automatic Dependent Surveillance-Broadcast) vehicles (estimating the rate of conflict with small Unmanned Aerial Systems (UAS) traffic is a subject in a separate large body of ongoing research).

Air risk has been studied in aviation for decades; there is no possibility to survey all the work. As far as air risk for drone operations is concerned, one of the following two technical alternatives was assumed in the different papers:

1) the conflicting traffic is distributed uniformly in the airspace, or

...
2) historical trajectory data for the conflicting traffic is available.

The first alternative (dubbed “Dutch model” in [15], due to its appearance in research led by Dutch [16]–[18], following the classical work of Weibel and Hansmann [19], is assumed in [3]–[6]. The second alternative was used in [7], [8] in which the aircraft trajectory data was taken from Eurocontrol’s NEST database and traffic density categorized as low, medium or high (the traffic density categories were fed into a belief propagation network outputting the conflict rate). The first approach may be more suitable for estimating conflicts with those vehicles whose positioning data is not available and thus the best guess is their probabilistic distribution over the space. The second approach can be applied to traffic with ADS-B or similar technologies onboard (cf. SORA being “limited to the risk of an encounter with manned aircraft” [1, C3.6]). In the remainder of this section we demonstrate such an application for Swedish lower airspace.

We split the region of interest into 58m-by-58m pixels (the pixel size defines the resulting map resolution, with lower pixel size the resolution becomes higher). We use OpenSky data [20] to calculate, for every pixel $p$, the number $n_p$ of aircraft that flew through $p$ or within 300m of $p$ during the time horizon $T = 16$hrs suggested in [7, p. 128] (Figure 1). The expected traffic intensity (number of flights per unit time) in $p$ is then $n_p/T$; e.g., if 16 aircraft flew through the pixel, then the intensity is $16\text{aircraft}/16\text{hrs} = 1\text{aircraft/hr}$. We assume that the intensity is constant during $T$; for finer-grained risk evaluation, a shorter $T$ (morning, day, night, specific hour) can be taken. Let $t$ denote the time it takes the drone to pass through a pixel; we assume $t$ is constant, independent of how the drone goes through the pixel, but, again, our analysis extends straightforwardly to arbitrary speed profile of the drone mission. The expected number of conflicts in $p$ is then $n_pt/T$, and the total expected number of conflicting aircraft along a drone path $P$ is the sum

$$\sum_{p\in P} \frac{n_p t}{T}$$

over all pixels in the path. To calculate the per-hour rate, divide that number by the time it takes to fly the path, which is $|P|/t$ where $|P|$ is the number of pixels through which the path goes. The conflict rate is thus equal to

$$\frac{\sum_{p\in P} n_p}{|P|/T} = \frac{\text{mean}(n)}{T}$$

– the average of $n_p$ over the path’s pixels and time.

Finally, to calculate the EFR, we follow SORA’s “assumptions on UAS lethality” [1, C3.7] and multiply the conflict rate by the number of people onboard aircraft – here again we parallel SORA which “does not consider the ability of the threat aircraft to remain well clear from or to avoid collisions with the UAS” [1, C3.5]. The air risk estimations in [7, p. 130] used 180 people, on average, onboard Boeing-737 or Airbus-320. From OpenSky data (Fig. 2) it can be seen that Stockholm Arlanda’s fleet is dominated by comparable fleet types; therefore we multiply the conflict rate by 180 for our EFR estimates. (Even though in general the number of passengers onboard may not be known with certainty, for finer-grain analysis, one may make separate estimates for different fleet types.)

Figure 3 left shows the EFR heatmap for direct Unmanned Aerial Vehicles (UAV) flights starting at a hospital in Stockholm. It can be seen that from HFRM point of view, the air risk is either prohibitively high or negligible, depending on whether the flightpath intersects aircraft routes or not. In the former case, SORA becomes crucial, with its mitigations involving, in particular, interaction with the air navigation service provider (ANSP). Note that in addition to Arlanda airport (the major Swedish international hub), Stockholm Terminal Manoeuvring Area (TMA) hosts also the smaller Bromma airport. The air risk map helps to pinpoint that it is Bromma which is the main contributor to the air risk for drone operations in Stockholm metropolitan area. Figure 3 right shows the EFR heatmap with only Arlanda flights (excluding Bromma). Comparing Figure 3...
left and right, we see that if Bromma traffic is removed, significant part of Stockholm airspace is air-risk-free for drone missions.

Stockholm TMA has busiest traffic in Sweden; in other places of the country, the contribution of air risk to EFR is much smaller. For example, Figure 4 shows the occupancy of pixels in the Swedish municipality (≈ 100000 inhabitants) of Norrköping (analogously to Fig. 1, showing it for Stockholm). The figure also shows a set of drone paths in Norrköping (we consider the ground risk for the paths in the next section). For all these paths, the air risk EFR is virtually zero because the considered UAV operations and the manned aviation operations are well separated by altitude. For such scenarios, ground risk is essentially the sole contributor to EFR – we turn to considering the ground risk in the next section.

III. GROUND RISK

Studying ground risk is a more recent development, motivated, in particular, by Very Low Level (VLL) Beyond Visual Line of Sight (BVLOS) operations close to the population on the ground; see [21] for the survey. In all prior work [2], [3], [9], [22], the ground risk was handled with pixel-based algorithms: for every pixel in the domain, the algorithms calculated how many people would be affected if the drone fails in the pixel, and then the grid of pixels was searched for least-cost path, where the cost is a linear combination of the path length and number of people affected in the pixels through which the path goes. Since grid-based least-cost routing may be suboptimal (Fig. 5), here we use the algorithm from [12] which uses paths with unrestricted orientation of edges and outputs Pareto-optimal paths, i.e., shortest paths with given risk and lowest-risk paths for a given length (users may also produce their own paths in the GUI tool https://undefined.github.io/ground_risk/ accompanying [12]). The algorithm is edge-based (instead of pixel-based): the number of affected people is calculated for every edge of the graph on pixels, in which edges have arbitrary orientations and are not restricted to connect only neighboring pixels (as was the case in earlier, pixel-based algorithms).

Figure 6 shows an application of the algorithm in [12] to the Swedish municipality of Norrköping. The heatmap is the population density from [23]. Getting real-world data on density of population, potentially impacted by drone failure, is a major open question in UTM (the most common hope is to use mobile phone density). In the absence of a reliable source, census statistics is often used for the population density [8], [12], [24]–[27]. Our algorithms may run on arbitrary density data, and for the illustrative examples presented here, census data serves the purpose.

Each of the routes in Fig. 6 left is a Pareto-optimal path, i.e., it has minimum risk for its length and is the shortest path among paths with the same EFR; Fig. 6 right shows the lengths and EFRs of the paths. Note that not all paths have EFR below 10^{-6}, and thus not all of them would be permitted from HFRM point of view.

IV. QUALITATIVE RISK ASSESSMENT: A FLAW, AND A FIX FROM HFRM

Examples like in Figure 6 elucidate an intrinsic drawback of a 0/1 (forbidden/permitted) view on risk assessment, oblivious to careful quantification of the risks. For instance, SORA would assign the same GRC to all routes in Fig. 6 (In this example GRC would be 6 for all routes, because the initial and final parts of the flights are in densely populated areas; if an operator would claim that these parts are short and may be neglected, we could scale up the example so that the parts become longer, up to the point where the operator would agree with the GRC 6.) After applying mitigations to reduce Specific Assurance and Integrity Levels (SAIL) to an acceptable number, the operator will have no motivation to fly the route with smaller EFR – the operation is considered safe enough (risk is under the threshold), and further decrease of the risk does not immediately “buy” anything for the operator. This is despite the fact that the EFRs imposed by the routes in Fig. 6 differ 6-fold, ranging from < 0.5 × 10^{-6} for the longest path that goes around the populated area of the city to > 3 × 10^{-6} for the shortest path which is simply a straightline segment between the two endpoints.

Our concern, expressed in the paragraph above, echoes CORUS which noted how ARC discourages attempts to decrease GRC: “drone operations in ARC-c or ARC-d offer little motivation for an operator to reduce the GRC, as the SAIL stays almost unaffected.” [28] Section 2.2.2.c. Here we note how GRC similarly discourages attempts to decrease EFR.

To fix the above flaw of the 0/1 risk assessment view, we suggest mandating the use of quantitative risk assessment (in addition to SORA and similar methods) for operation approval. If our suggestion is followed, an operator would be required (or encouraged, at the initial implementation stages) to use a longer but safer route (with low EFR), among the SORA-approved routes. Such a requirement would encourage the operator to employ computerized tools, like the one used in this paper, for mission planning and approval. This is in line with the emphasis on automation and digitalization in UTM, included in the very definition of U-space service as one “relying on digital services and automation of functions designed to support safe, secure and efficient access to U-space airspace for a large number of UAS” [29] Article 2(2)]. Indeed, without mandatory usage enforced by the regulators, automation and digitalization risk being empty hype words in the definitions, while risk assessment and operation approval is done with pen and paper. At the early stages, the opportunity to get operation approval without using a computer may be seen as an advantage of qualitative methods; however, the future of the industry will likely rely on computerized services which optimize the performance and minimize the adverse impacts of drone proliferation (see e.g., CORUS [28] for the many suggestions of such services expected in U-space).

As a last remark, we again draw attention to CORUS saying (in the same section that noted ARC domination and the implied lack of motivation for operators to reduce GRC) that...
“GRC can be reduced easily using the information provided by certain U-space services.” [28, p. 11]. The tools used in this paper for producing Pareto-optimal paths may contribute to such services, specifically targeted to address ground risk issues.

V. HF RM: A FLAW AND A FIX FROM PARETO

In the previous section we advocated for a more aggressive introduction of digitalization and automation on the example of supplementing qualitative risk assessment with HFRM. In this section we show that HFRM is not a panacea: applied blindly, it may issue a “license to ignore fatalities” beating the purpose of risk management.

Indeed, technically speaking, EFR of a mission may be lowered by the following trick (Fig. 7): the drone spends a + R √ 2 + 1 time in a low-risk area and then plows through a high population density (which is not an assembly of people, so flying over it is allowed). Thanks to the wasteful part of the path, its EFR will be low because for a fixed number of exposed people, EFR is inversely proportional to the path length (for a formal proof of this intuitively clear observation see, e.g., [12, Eq. (2)]).

One way to deal with the above kind of “cheating” could be to forbid unduly long paths. But how to define “unduly long”? Pareto optimality comes to rescue again: the regulator may mandate that drones use only Pareto-optimal paths, i.e., only shortest paths with a given risk.

To deny permit for an unduly long path, the regulator may
employ tools similar to those we used in Section III. As mentioned in the previous section, the operators may use such tools too. In the end, it does not matter whether an operator deliberately extends the path to decrease EFR or makes an “honest mistake” of computing a suboptimal path (due to absence of the necessary mission planning tools); since social acceptance of UAV operations is crucial for the drone industry growth, it is important that the drone flights annoy the citizens as little as possible, or, in our terms, do not fly more than it is necessary for avoiding high risk areas. Hence we believe that in the latter case (an operator non-intentionally losing efficiency due to an unduly long path), the permit denial is justified. After all, performance-based services (PBS) are adopted in the conventional aviation and are not an equity breach; on the contrary, favoring those who invest into better equipment and advanced computational tools serves as a driver for the industry. Last but not least, the (Pareto) optimality mandate may selectively apply only to certain class of operations: drone operations may be classified into “Loitering” and “Delivery”, with only the latter being subject to the “optimal” regulations.

Robust statistics

We finish this section with a suggestion of a purely technical way to fix the HFRM flaw described above: use the median fatality rate, instead of the expected fatality rate (EFR), as the measure of the path’s risk. Median is generally known as being more robust than the mean (expected value) – and this is the property of the median on which we want to capitalize.

Specifically, consider the graph of the risk level as function of the location along the path (Fig. 8). Then the expected (or mean, or average) risk of the path is the height of the rectangle that has the same area as the integral of the risk along the path (the area under the graph). The median risk is the height that has the same under-the-graph area below it as above it (i.e., the “risk mass” below and above the median is the same, or in other words, half of the area under the graph is below the median and half is above). In particular, if the risk of a path is measured as its median risk, paths with high concentration of risk (such as ones on Fig. 7) are rightly assigned high risk, irrespective of the “cheating” that a path may do by “gaining” (wasting) length in a low-risk region: note from Fig. 8 that the median level of the blue path remains the same as for the red path – the height that cuts half of the area under the graph is the same for the red and the blue graphs.

We admit that changing from expected to median risk is paradigm-shifting (long-term average risk rates have been used in aviation for long time) and may deserve further investigation.
This paper contributes to understanding of interplay between qualitative and quantitative methods for risk assessment. We highlighted intrinsic flaws in both methodologies and suggested ways to fix the flaws. We also argued in favor of regulations enforcing (Pareto-)optimality of drone operations, based on quantitative risk assessment tools.

Naturally, we are not the first to suggest increased levels of automation in risk assessment (the EU ATM Master Plan [30, p. 24] defines 5 levels of automation, simplified from the more extensive taxonomy described in SESAR’s Automation roadmap). Suggestions to digitize/automate SORA can be found in GUTMA’s (Global UTM association) Annual conference announcement [31] which had “Defining how UTM services can best support the digitalization of airspace risk assessment (SORA)” as one of “the immediate hard challenges facing UTM”, as well as in Riga airspace assessment [32] which noted: “However, the requirement to assign ground risk classes for UAS operation has prompted the conclusion that this process needs to be digitalised and automated to make significant resource savings for UAS operators and authorities, at least for determining the intrinsic UAS Ground Risk Class (GRC).” Still, it may be hard to automate SORA itself, given SORA’s “room for interpretation” [3]. For example, SORA says that “In case of a mismatch between the maximum UAS characteristic dimension and the typical kinetic energy expected, the applicant should provide substantiation for the chosen column.” [1] without specifying what the “substantiation” may be, which let e.g., [3] choose the first column when defining the GRC for a drone with wingspan > 1m but kinetic energy < 700J. An alternative way is to have automation complement SORA – the approach taken in this paper.

A lot of work remains before the best way to combine the qualitative and quantitative methodologies is found. For example, on the technical frontier, it may be interesting to “EFR-quantify” SORA’s mitigations, i.e., give algorithms to quantify how the mitigations decrease the risk [33]. Similar studies may be done on other risk assessment methods, in particular on those practiced outside EU. On the regulatory side, further investigations are needed in order to accept specific computational tools as mandatory for drone missions planning and approval.

ACKNOWLEDGMENTS

This research is partially supported by the Swedish Transport Administration, the Swedish Research Council, CORUS-XUAM project which has received funding from the SESAR Joint Undertaking under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101017682, and AirMOUR project which has received funding from the EU’s Horizon 2020 research and innovation programme under grant agreement 101006601. The work by TM and MU was done during their stays with Linköping University. We also acknowledge the anonymous referees whose comments helped with the presentation of the paper.

REFERENCES


