A Comprehensive Evaluation of U-space KPIs

Juan Nunez-Portillo
Department of Aerospace Engineering
Universidad de Sevilla
Seville, Spain
jnportillo@us.es

Jonas Lundberg, Tatiana Polishchuk
Valentin Polishchuk, Leonid Sedov
Communications and Transport Systems
Linköping University (LiU), Sweden
firstname.lastname@liu.se

Gabriele Enea
MIT Lincoln Laboratory
ATC & Weather Systems Group
Lexington, USA
gabriele.enea@ll.mit.edu

***Abstract*—EASA's Acceptable Means of Compliance and Guidance Material (AMC/GM) to Regulation (EU) 2021/664 on a regulatory framework for the U-space put into writing the obvious fact (obvious to U-space stakeholders) that U-space airspaces must be regularly assessed based on a well worked out set of performance metrics. This paper reports on a probabilistic analysis of a variety of Key Performance Indicators (KPIs) for UAV (Unmanned Aerial Vehicle) Traffic Management (UTM). The KPIs belong to the Key Performance Areas (KPAs) of Access and Equity, Capacity, Efficiency, and Safety. The metrics are evaluated on a variety of simulated demand scenarios in very low level (VLL) metropolitan airspace over European cities. In particular, we study the dependence of the UTM KPIs on the lookahead time of strategic deconfliction of U-plans; the length of this time interval, between the deconfliction and the desired takeoff time of the drone operation, is called Reasonable Time to Act (RTTA) in the European Concept of Operations for U-space (CORUS) and its Urban Air Mobility (UAM) extension (CORUS-XUAM). In view of the stochastic nature of VLL UAV traffic, we present not only the average values of the KPIs, but also their probability distributions.**

I. Introduction and Related Literature

Performance evaluation of air traffic services has been high on the agenda of ATM modernization efforts both within SESAR in Europe [1], [2] and NextGen in the U.S. [3] ("If you can't measure it, you can't improve it" [4]). The same applies to UTM – the ATM's counterpart for handling traffic of unmanned, a.k.a. uncrewed, aerial systems (UAS), a.k.a. drones, in very low level (VLL) airspaces: U-space regulatory framework underlines the need to establish and monitor a set of metrics for performance of U-space services. In particular, Article 18 of U-space Regulation (EU) 2021/664 [5] is a list of "Tasks of the competent authorities" related to the U-space; the Acceptable Means of Compliance and Guidance Material (AMC/GM) to this regulation [6] expand and clarify the tasks. For instance,

- GM2 to Article 18(f) of the regulation reads
  "The complexity of the U-space airspace should be addressed through its different life cycle phases".
- GM2 to Article 18(f) of the regulation further suggests
  "establishment of indicators/KPIs or metrics (covering aviation performance, safety metrics, and sustainable urban mobility) at the national, regional and local level of the U-space deployment".

This paper shows what can be done to comply with the regulations on a local level; future work may address the same questions on the regional and national levels.

In addition to the above quotations from AMC/GM,

- GM1(g) to Article 3 "U-space airspace" reads
  "The regular reassessment of the U-space airspace is expected to be conducted by the Member States to evaluate its effectiveness in supporting the safe, secure, and efficient conduct of UAS operations."
- GM1 to Article 18(g) says
  "The competent authority may monitor or audit authorisation and rejection data to assure equitable access to airspace" in order to support operational performance.

One way the flight authorization service can contribute to more efficient and equitable airspace access [7]–[13] is by departing from the deconfliction of U-plans on the First-Come First-Served (FCFS) basis: AMC/GM [6, GM7 Art 3(4)] to Regulation 664 allow local authorities to define a time window during which flight activation may be requested, and to limit how long before the takeoff the flight authorization may be sought – these measures are envisioned to make the strategic deconfliction more efficient and fair to airspace users. The European Concept of Operations for U-space (CORUS) [14] and its UAM extension (CORUS-XUAM) [15] propose that instead of FCFS, U-plans may be deconflicted at a certain time before the desired start of the flight: the duration of this interval was dubbed RTTA (Reasonable Time to Act); in certain contexts [15, Section 2.2.3.2.1], RTTA also refers to the moment when the U-space Service Provider (USSP) checks whether the submitted U-plan can be executed as planned or should be modified to avoid conflicts. The overall idea with RTTA is that a U-plan reaching this point is considered "frozen", meaning it will not be modified unless a conflict arises with a priority flight (e.g., emergency or police operations). At RTTA, the USSP commits to allowing the flight to be executed as planned, assuming nominal operating conditions. On the contrary, a U-plan that has not yet reached its RTTA may still be rejected by the USSP (e.g., due to conflict with another flight), and a modification of the flight (such as delaying its start) may be suggested. This way, RTTA values balance early airspace reservation versus the flexibility of last-minute adjustments, addressing the following two conflicting

considerations:

- Deconflicting early (e.g., FCFS, as U-plans arrive) is unfair to operators who do not know their plans in advance (rapid response missions, on-demand services like delivery, breaking news hunters, etc.); early airspace reservation is also sensitive to uncertainties like cancellations for technical (drone or takeoff/landing area not ready), business (demand vanished), or any other reasons (canceled flights leave the reserved airspace unused).
- Deconflicting close to the planned flight start discourages early planning, and may disappoint an operator whose plan is rejected/delayed by the USSP close to the operation (thus not giving the operator a reasonable time to act upon the U-plan rejection); see more on this in DACUS project report [16, Section 3.2.1.3].

Note that U-plans submitted earlier than their planned start minus RTTA, are deconflicted in the order of their start times, while those submitted past their RTTAs are deconflicted in the order of their submission times. Also note that RTTA concept per se (as studied in this paper) does not involve any optimization (such as grouping flights into batches and optimizing the order within each batch w.r.t. some objective [17], [18]): we consider the most basic setup (without optimization) in order to explore RTTA impact on UTM KPIs.

A. Our contributions

In this paper, we assess UTM KPIs, for various RTTAs, in the areas of Access and Equity, Capacity, Efficiency, and Safety. Thus, among SESAR's 9 Key Performance Areas (KPAs) [1], we do not explicitly measure Predictability, Flexibility, Environment, and Security KPIs. While the Environment and Security KPAs are outside our scope, the RTTA notion per se is all about Predictability (when the situation at takeoff can be predicted confidently enough to freeze the U-plan?) and Flexibility (how long before the operation do drone operators have the flexibility to change their plans?) questions. Hence, overall, our work is related to 7 out of 9 SESAR's KPAs.

The specific KPIs measured are average delay (also per delayed drone) and maximum delay (Efficiency), number of delayed missions and number of drones delayed by more than 15 min (Capacity), Gini index (Equity), and number of potential conflicts (Safety). The KPIs are evaluated under a variety of scenarios: for each of the considered cities (Norrköping in Sweden and Ljubljana in Slovenia), the scenarios differ by the demand level (Low, Medium, High), traffic mix (scheduled vs. on-demand services), presence or absence of priority drones, and type of deconfliction (strategic vs. tactical). In addition to the baseline scenario without weather effects, we also consider 12 weather scenarios that differ in terms of forecast uncertainty (reflecting reduced reliability with earlier predictions) and wind field variability.

We also look at users diversity regarding their potential business needs: while all operators prefer to fly without delays and to know about delays early (i.e., having a large RTTA), different missions have different tolerances for the same delay D of the flight. For instance, inspections scheduled at regular intervals (e.g., weekly or monthly) may not care about being delayed for some time if they know about the delay well in advance and have the time to replan (have reasonable time to act); on the contrary, on-demand services, such as hot food delivery, do care about the delay and do not even know about the operation long in advance (thus, having a large RTTA is of no use for such an operation). We calculate the cost, or price $P(D, RTTA) = [D - c \cdot RTTA]_+$, in Airbus' terms [7], [9], of knowing about the delay D at the time $RTTA$ before the departure, where $[x]_+ = \max(0, x)$ (the price should not be negative, i.e., the operator does not get profit from knowing about the delay early). The value of c determines the relative importance of RTTA (having time to act after learning about the delay) vs. the cost of the delay: with $c = 0$ the user only cares about the delay; having larger c means that increasing RTTA (i.e., knowing about the delay earlier) decreases the operator's cost. We emphasize that the above formula for the price is used only for simplicity: our framework can handle any other formula for P, including non-linear dependence on the arguments (e.g., in [7], while the price depends linearly on D for "intermediate" delays, the price is non-linear for small and large delays).

We ran extensive simulations which allowed us to obtain not only the mean values of the KPIs, but also their distributions (our probabilities approach borrows methodology from SESAR's ER4 project FMP-Met). Parts of this paper build up on SESAR's Metropolis 2 [19] project (we use the same method for demand estimation; the project also evaluated UTM KPIs). Some probability distributions related to RTTA were explored also in the recent work by SPATIO project [10].

Roadmap: The next section presents the experimental setup, outlining connections to earlier literature on which our choices are based. Section III reports the results. We conclude in Section IV, including some directions for future research.

II. Simulation framework

We consider two urban scenarios: the municipalities of Norrköping (Sweden) and Ljubljana (Slovenia). Since they differ in size (1500 km^2 vs. 275 km^2, resp.), we adopt 12×12 km^2 squares centered at the respective city centers (58.5885°N, 16.1883°E and 46.0514°N, 14.5061°E), similarly to the earlier work [7], [9], [20]–[25]. Following the approach in the Metropolis 2 project [19], we apply the method from [26] to set UTM demand levels: the method estimates drone-based package delivery volumes in European regions based on a set of inputs, which are presented in Table I for our use cases (the table also shows our outputs). Guessing that parcel delivery will constitute approximately three fourths of overall UAV activity, we derive rough estimates of hourly U-space traffic demand to be 600-750 requests for Norrköping and 950-1250 for Ljubljana. For comparability, we adopt a common set of demand levels for both cities, by selecting the maximum estimated value for each city, along with an overconservative upper bound of 1750 requests/hr: the resulting demand levels are 750 (low), 1250 (medium), and 1750 (high). We remark

that understanding the demand for drone services is a million-dollar question that we are by no means trying to answer here; our experiments merely demonstrate the difference in KPIs for various demand levels.

A. Model features and parameter configuration

We simulated drone traffic using the Cal model [9], [20]–[24] (named "Cal" in [25] after its introduction by UC Berkeley at ICRAT 2016 [28]); the model is an extension of the probabilistic setup from Hoekstra's PhD thesis [29], further developed by Jardin [30], and later explored within the Metropolis project [31]–[34] (such model is referred to as "Dutch model" in [35]). In both the original Dutch and the Cal model, traffic is two-dimensional (thus precluding using altitude for strategic deconfliction), with drones flying direct from origin to destination at a constant altitude (see SESAR's ICARUS project [36] and [37] for studies of common altitude reference for UTM, unifying AGL and AMSL height measurements). The Cal model generates origin-destination pairs based on population density (similarly to, e.g., the Metropolis 2 project [38], [39]); we used 100-meter resolution population data from the Global Human Settlement Layer (which we used also in the demand estimation, see Table I, footnote e).

As in [7], [9], for the baseline scenario we used the safety margin $r = 150$m (drones are in conflict if they are at horizontal distance less than r) and the drone speed of $v = 25$m/s. As in [7], [9], [24], we apply ground delay for deconfliction: a flight is delayed until it no longer conflicts with earlier scheduled flights, and then just flies along its planned route. Following the RTTA principles outlined in Section I, any flights past their RTTAs are frozen (flying as scheduled). For example, if a flight C-D is delayed because it conflicts with a flight A-B, then flight C-D will still depart delayed even if flight A-B is later canceled – again, the logic is that after RTTA the operator prefers to have no changes (even if now a change would be favorable) because the operator does not have a reasonable time to act on the change (update the delivery time, reschedule charging, etc.).

In our discrete-event simulation, system evolution is represented as a sequence of events: U-plan submission, conflict detection and resolution, cancellation, departure, landing. The inherent variability of U-space operations is captured by randomness in the origin and destination selection (sampled proportionally to the population density), U-plan submission times, requested takeoff times, and cancellations (as in [9], the probability of cancellation is a decreasing linear function of the file-ahead time, falling from 10% at 24hrs to 0% at 10min before the start).

Modeling weather impact is done in two ways, both taken from [9]: by varying the speed v and the safety radius r for the drones. The speed variability was due to taking into account the wind effects: estimating the expected wind-informed speed shows that stronger winds lead to decreased ground speed v; we experimented with the speeds in the range $v \in \{15, 20, 25\}$m/s. The changes in r reflect the fact that longer weather forecast lookahead times implies larger uncertainties which lead to increased safety margins: we considered $r \in \{150, 175, 200, 225\}$m. We emphasize that we do not postulate any specific dependence of r on the lookahead time (this is outside our expertise and the scope of this research): our methodology works with any dependence because the input to our framework is just the radius. If a relationship between the radius and the lookahead time is developed in the future, this could be plugged into the model. Progress towards that end are currently being worked and will be published in [40].

In some simulations we set 10% or 20% of the drones to execute priority missions, meaning they cannot be delayed – they fly when they need to. We consider two scenarios:

- *Strategic deconfliction with priority missions*. Any regular drone must be delayed if it conflicts with priority operations. In our framework, this is easily modeled by assigning an infinite RTTA to all priority drones (we denote their RTTA by $RTTA^*$). This ensures that by the time any non-priority U-plan is being deconflicted, all priority drones have already been assigned their requested 4D volumes. In effect, priority missions reduce the available airspace capacity for the other operations.
- *Tactical deconfliction with priority missions*. The priority are on-demand operations, i.e., they are not known in advance: they just pop up and fly. In our framework, this is easily modeled by assigning a zero RTTA to all priority drones ($RTTA^* = 0$). In particular, if a regular drone is airborne when a priority mission pops up and flies, it is too late to delay the regular drone.

Note that we do not deconflict priority vs. priority, assuming that this is done outside the "normal" UTM, similar to how civil air traffic controllers only guide commercial aircraft around military operations, while separation between military flights is delegated to military ATCOs.

B. Initialization

To obtain statistically meaningful estimations of every KPI and characterize its distribution, we need to gather statistics from multiple independent simulation runs, each initialized with a different random seed end each being sufficiently long [41, Chapters 1, 4, 9]. Instead of deciding beforehand the number of runs (as was done in previous similar studies, e.g., 3000 runs in [11], 120-155 in [42], and 1000 runs in [43]), we continue with the runs until a stopping criterion, based on statistical convergence, is met. Because we are interested in the full distribution, and tail estimates require more samples than central tendencies (mean/median), we focus on the 95th percentile, q_{95}, and use percentile bootstrap [44] to build a confidence interval (CI) for $\hat{q}_{95}$: we terminate the simulations when the CI width relative to the estimate size $|\hat{q}_{95}|$ falls below a threshold of $\epsilon = 5\%$ [45], i.e., CI width $< \epsilon \cdot |\hat{q}_{95}|$.As for the length of one run, it was acknowledged in [7], [43] that the system exhibits a warm-up period that must be discarded: to understand these effects, we simulated 24hrs of traffic for several RTTAs. Figure 1 (top) illustrates a sequence of hourly mean delays, with an initial warm-up period characterized

TABLE I
INPUTS AND OUTPUTS FOR PARCEL DELIVERY DEMAND ESTIMATES FOR 2035, BASED ON THE EU ATM MASTER PLAN [27].

City	Annual parcel volume (country level)	% living in urban areas (country level)	Economic growth rate (%, country level)	% living in the area of interest	Hourly demand (pessimistic)	Hourly demand (realistic)	Hourly demand (optimistic)
Norrköping	202M[a] (2022)	89[c] (2023)	1.1[d] (2025)	1.1[e] (2025)	465 (2035)	492 (2035)	548 (2035)
Ljubljana	40M[b] (2024)	56[c] (2023)	2[d] (2025)	13[e] (2025)	714 (2035)	788 (2035)	957 (2035)

[a] https://www.pitneybowes.com/content/dam/pitneybowes/us/en/shipping-index/23-mktc-03596-2023_global_parcel_shipping_index_ebook-web.pdf
[b] https://www.statista.com/statistics/1219822/courier-express-parcel-market-volume-slovenia/
[c] https://data.worldbank.org/country
[d] https://economy-finance.ec.europa.eu/economic-surveillance-eu-economies_en
[e] https://human-settlement.emergency.copernicus.eu/download.php?ds=pop

by lower delay values and smaller differences across RTTA settings (for visualization clarity, high-frequency oscillations in hourly delays are smoothed out by applying a moving window over averaged processes [46]; it can also be seen that the system exhibits a cut-off effect near the end of the simulation, which must also be discarded. As a result, each simulation is run for a total duration that includes the warm-up period plus two hours, and only the middle hour is retained for KPI computation, with the warm-up and final hour both discarded.

Interestingly, under certain setups described in Section II-A, the system does not reach steady-state conditions (Fig. 1 (bottom)): delays accumulate over time, indicating that the demand exceeds the airspace capacity. However, since real-world operations follow a diurnal traffic pattern with natural peaks and valleys, such unstable regimes are less concerning in practice. To reflect this, we assume the daily operations last for 8hrs and conservatively select the results from the last, i.e., the 8th, hour. For simplicity and consistency, we harmonize the methodology across both stable and unstable regimes by always running simulations for nine hours, selecting the results from the 8th hour: since in stable settings, the system reaches steady-state conditions before 8hrs, selecting a later hour changes nothing; in unstable settings, we are on the safe side by selecting a late hour, but still reflecting conditions compatible with normal operations.

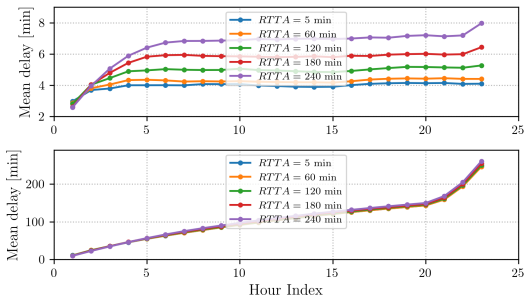

Fig. 1. Illustration of stable (top) and unstable (bottom) regimes for Norrköping with safety radius $r = 200$ m, for various RTTA values. The top panel shows results averaged over multiple long runs at 750 requests/hour, and the bottom panel shows a single long-run outcome at 1750 requests/hour.

III. RESULTS

This section presents a selection of results: a comparison between the 2 cities across 3 demand levels, an evaluation under various r and v combinations (modeling the weather impact), and an analysis of different priority mission percentage.

A. Baseline scenarios

We computed the distribution of each KPI as a function of RTTA for both Ljubljana and Norrköping at the demand levels of 750, 1250 and 1750requests/hr; as illustrative examples, Figures 2, 3 and 4 present the results for mean delays, the number of drones delayed by more than 15 minutes, and mean price respectively. The overall system performance, as indicated by the mean delay (Fig. 2), deteriorates with increasing RTTA, regardless of city or demand level (small RTTA creates additional capacity leading to smaller delays); the deterioration becomes more significant at higher demands, and remains mostly similar between the cities, with the most noticeable divergence at 1250requests/hr. The KPIs distributions are symmetric (this finding is valid for all scenarios considered in this study), and their dispersion (i.e., difference between 95th and 5th percentiles) slightly grows for 750 and 1250requests/hr, but remains constant at the high demand. As an example application of the distribution, if RTTA = 120min at 1250requests/hr, then the average delay is $<$ 5min with probability 30% for Ljubljana and 69% for Norrköping. Interestingly, at 750requests/hr, Norrköping is more affected by the delays than Ljubljana, while at 1750requests/hr, Ljubljana is more impacted. Although the number of delayed drones is consistently higher in Norrköping (not shown due to space constraints), the number of drones delayed by more than 15 minutes (Fig. 3) increases more rapidly for Ljubljana as demand rises; at 1750requests/hr, this number is around 60% higher in Ljubljana compared to Norrköping.

As explained in Section I, the cost, or price, of being informed about the delay at the time RTTA before the departure is not the same for all operators. We assigned the values of $c = 0$ or $c = 0.5$ to the drones uniformly at random; recall that $c = 0$ prioritizes delay minimization (e.g., on-demand services), while $c = 0.5$ models operators for whom anticipation is also important (e.g., missions planned well in advance). The resulting trends are presented in Fig. 4. It can be seen that in the extreme cases, where all drones are assigned either $c = 0$ (dotted curve) or $c = 0.5$ (dashed curve), the

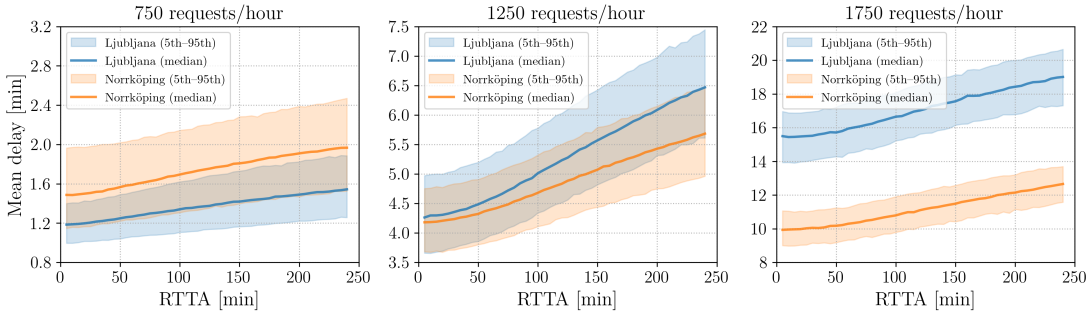

Fig. 2. Mean delays as a function of RTTA for Ljubljana and Norrköping at 3 demand levels, with $r = 150$m and $v = 25$m/s.

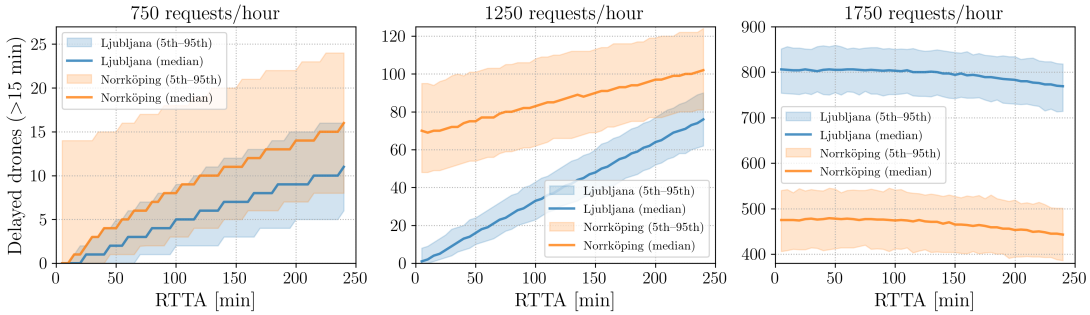

Fig. 3. Number of drones delayed by $>$ 15min as a function of RTTA for Ljubljana and Norrköping at 3 demand levels, with $r = 150$m and $v = 25$m/s.

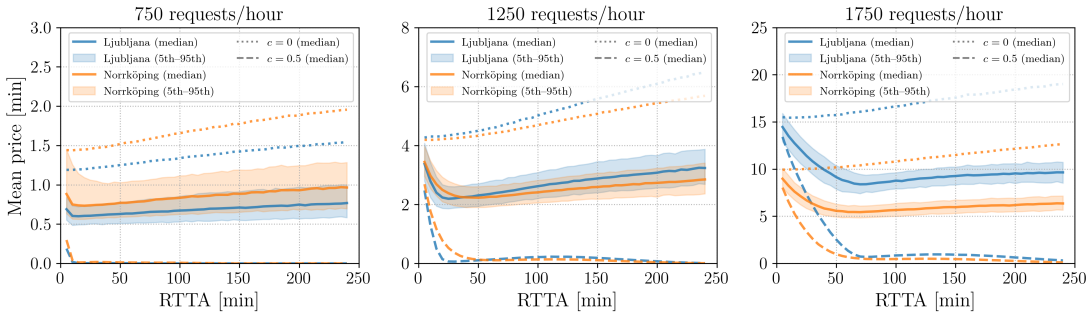

Fig. 4. Mean price as a function of RTTA for Ljubljana and Norrköping at 3 demand levels, with $r = 150$m and $v = 25$m/s. Drones are have $c = 0$ or $c = 1/2$ with probability $1/2$. Median results for $c = 0$ and $c = 0.5$ are also shown.

mean price increases or decreases with RTTA, respectively. When both cases appear in the traffic mix, the resulting price evolution exhibits a minimum at low RTTA values; the position of this minimum shifts rightward as the demand increases.

B. Impact of weather

We computed the distribution of each KPI as a function of RTTA for Ljubljana at 750 requests/hr for $v = 25$m/s (baseline), 20m/s and 15m/s. For each v and RTTA, safety radii $r \in \{150, 175, 200, 225\}$m, accounting for scenarios of different levels of weather forecast uncertainty, were considered. As illustrative examples, we present results for the number of delayed drones (Fig. 5) and maximum delays (Fig. 6).

As can be seen from Fig. 5, the number of delayed drones is already high under the baseline scenario ($v = 25$m/s and $r = 150$m). For instance, looking at the raw data, if RTTA = 5min, at least 75% of post-RTTA U-plans will need to be modified with probability 86%. However, most of these delays are relatively small (not shown due to space constraints), which suggests using flexible departure windows rather than rigid departure times, as proposed in [11]. While the number of delayed drones decreases with RTTA, the severity of their delays rises (Fig. 6); the maximum delays also increase with stronger winds – decreasing speeds effectively "slow" the time and more conflicts are created as drones spend more time in the air, flying to their destinations. Moreover, the dispersion in the maximum delays gets wider with increasing RTTA, ranging from 4min at RTTA = 5min to 66min at RTTA = 240min.

As far as the different combinations of r and v are concerned, several findings can be highlighted:

- As v decreases, the number of delayed drones becomes

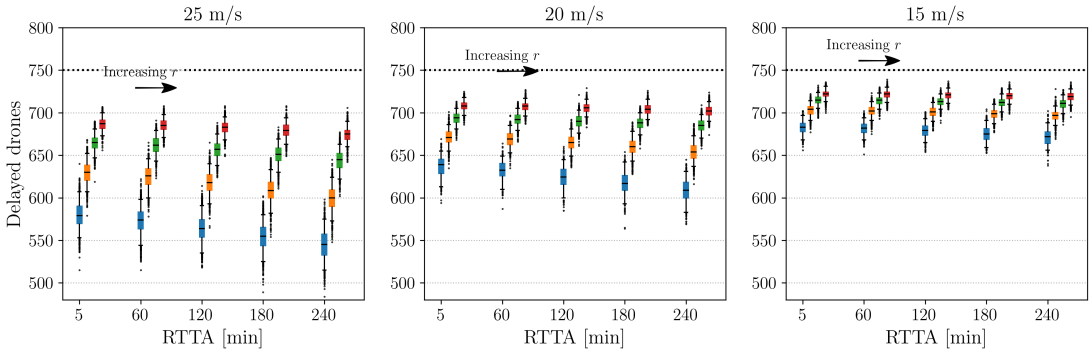

Fig. 5. Number of delayed drones as a function of RTTA for Ljubljana at 750requests/hr, evaluated across 3 groundspeeds. For each combination of groundspeed and RTTA, results are shown for $r = 150, 175, 200, 225$m with blue, orange, green, red resp.

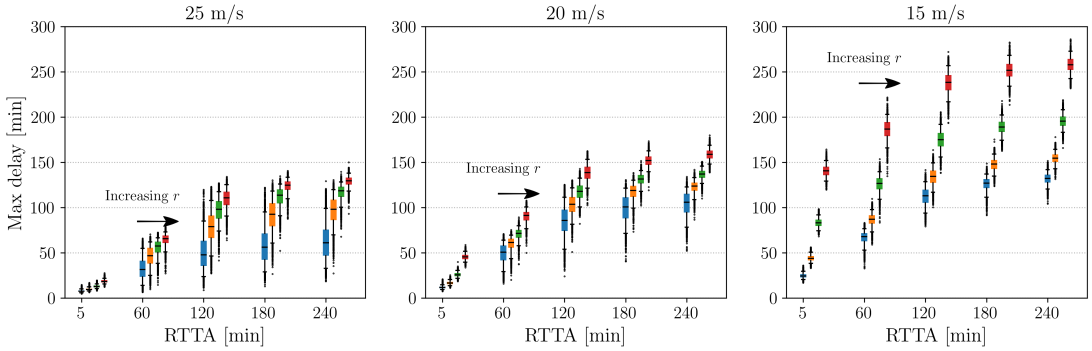

Fig. 6. Maximum delay as a function of RTTA for Ljubljana at 750requests/hr, evaluated across 3 groundspeeds. For each combination of groundspeed and RTTA, results are shown for $r = 150, 175, 200, 225$m with blue, orange, green, red resp.

nearly independent of RTTA, while its influence on maximum delays grows.
- As r increases, both the number of delayed drones and maximum delays increase.
- For maximum delays in particular, as either v decreases or r increases, the dispersion no longer varies monotonically, but instead exhibits a peak as function of RTTA.
- For the number of delayed drones and for all v, as r increases, the dispersion decreases.
- For maximum delays, as v decreases, the maximum dispersion shifts toward higher values of r.

C. Baseline scenario with priority flights

We computed the distribution of each KPIs as a function of RTTA for Ljubljana at 1250requests/hr, across 3 scenarios: 0%, 10% and 20% of drones designated as priority, with $RTTA^* = 0$ or $RTTA^* = \infty$. As illustrative examples, we present results for the number of conflicts with priority drones (Fig. 7), mean delay per delayed drone (Fig. 8) and Gini index (Fig. 9).

For a given proportion of priority drones, the number of conflicts between regular and priority drones (Fig. 7) is generally lower when $RTTA^* = 0$ than when $RTTA = \infty$. This is expected: when $RTTA^* = \infty$, conflicts can occur with all regular drones, whereas when $RTTA^* = 0$, they can only occur with regular drones still on the ground (see end of Section II-A). In addition, the number of conflicts shows little dependence on RTTA for $RTTA^* = \infty$, but decreases very fast with RTTA for $RTTA^* = 0$. Regarding variability, dispersion for $RTTA^* = \infty$ remains constant, whereas it exhibits a peak at some intermediate RTTA for $RTTA^* = 0$.

For the realistic low proportions of priority drones considered in this study, the KPIs follow nearly monotone trends that are easier to describe, though not entirely straightforward. As the proportion of priority drones increases, conflicts rise in both cases at low RTTA values, but for $RTTA^* = 0$ they show a slight decrease at higher RTTA values. Notably, as the proportion of priority missions increases, dispersion decreases.

As for the mean delay per delayed drone (Fig. 8), the case with $RTTA^* = \infty$ follows a similar trend to the baseline scenario, but with slightly higher delays (which increase with the proportion of priority drones). In contrast, when $RTTA^* = 0$, delays follow a logarithmic-like trend over RTTA, with lower (and less dispersed) delays at very small RTTA and higher (and more dispersed) delays at large RTTA.

Regarding equity, we evaluated Gini index (0 is perfect equality, 1 is worst inequality). The same observations made in the previous paragraph apply here, but with narrower dispersion (Fig. 9). Consequently, a low RTTA is preferable from the fairness perspective, particularly when $RTTA^* = 0$.

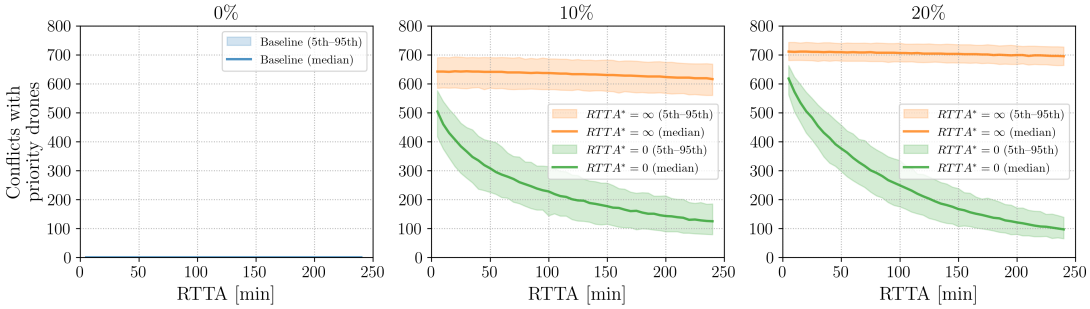

Fig. 7. Number of conflicts with priority drones as function of RTTA for Ljubljana at 1250requests/hr, evaluated across 3 scenarios with 0%, 10%, and 20% of drones designated as priority. Results are shown for both $RTTA^* = 0$ and $RTTA^* = \infty$.

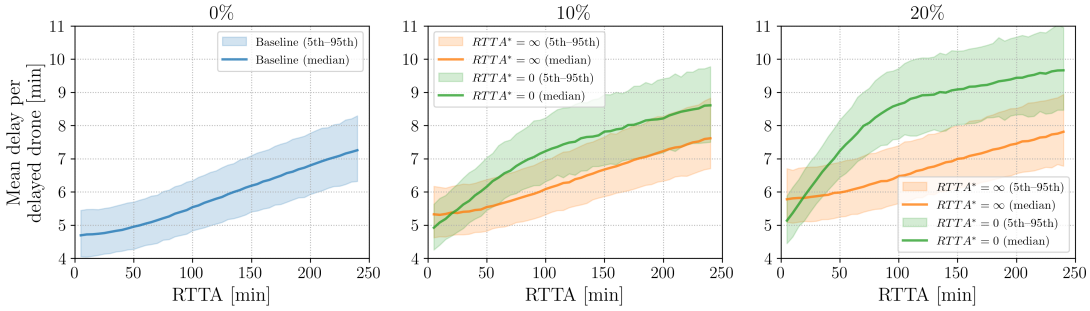

Fig. 8. Mean delay per delayed drone as function of RTTA for Ljubljana at 1250requests/hr, evaluated across 3 scenarios with 0%, 10%, and 20% of drones designated as priority. Results are shown for both $RTTA^* = 0$ and $RTTA^* = \infty$.

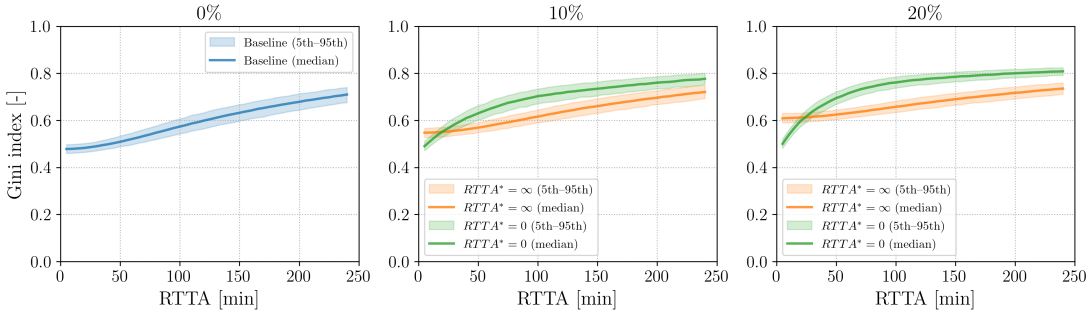

Fig. 9. Gini index as function of RTTA for Ljubljana at 1250requests/hr, evaluated across 3 scenarios with 0%, 10%, and 20% of drones designated as priority. Results are shown for both $RTTA^* = 0$ and $RTTA^* = \infty$.

IV. Conclusion

We evaluated a set of UTM KPIs on a variety of simulated scenarios for drone operations in two European cities. The evaluation produced not only the mean values, but also the distributions of the KPIs. Studies like ours (KPIs and the U-space assessment in general) will contribute to addressing the requirements laid out in EASA's AMC/GM [6, GM2] to Article 18(f) of EU U-space Regulation 664 [5] (planning and review phases, in particular – the technical review). It will allow competent authorities to make an important technical step towards compliance with the regulations and performance-based (instead of prescription-based) regulations overall. Future work may consider other costs/profits associated with the delays and RTTA, look at more KPIs, and perform simulations in more cities (even though most urban areas tend to have similar demand distribution, with heavily populated city center and sparser residential suburban areas, it may be interesting to see the results specifics), using various other scenarios and demand generation methods (e.g., [47]), including various flight cancellation models.

Acknowledgments: We thank anonymous reviewers for their helpful comments. This research is partially supported by the Swedish Transport Administration, Norrköping municipality, Vinnova project Testbed Air Mobility (grant dnr 2023-03716), the R&D&i project PID2021-122323OB-C32 (funded by MCIN/AEI/10.13039/501100011033/FEDER, UE), and Catalyst funding – a project supported by the SESAR 3 Joint Undertaking and its founding members under the Grant

Agreement nr. 101114648. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the EU or SESAR 3 JU. Neither the European Union nor the SESAR 3 JU can be held responsible for them.

REFERENCES

[1] SESAR, "PJ19. 04: Performance Framework," 2019.

[2] X. Prats, C. Barrado, A. Vidosavljevic, D. Delahaye, F. Netjasov, and D. Crnogorac, "Assessing ATM performance with simulation and optimisation tools the apache project," in *7th SIDs*, 2017.

[3] FAA, "Report on NextGen Performance Metrics," FAA, Tech. Rep., 2021.

[4] S. Ratcliffe, "Lord Kelvin," 2016. [Online]. Available: https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236

[5] European Commission, "Commission Implementing Regulation 2021/664 on a regulatory framework for the U-space," 2021.

[6] EASA, "AMC/GM to Implementing Regulation (EU) 2021/664," 2021.

[7] A. D. Evans, M. Egorov, and S. Munn, "Fairness in Decentralized Strategic Deconfliction in UTM," in *AIAA Scitech 2020 Forum*, 2020.

[8] C. Chin, K. Gopalakrishnan, M. Egorov, A. Evans, and H. Balakrishnan, "Efficiency and Fairness in Unmanned Air Traffic Flow Management," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 9, pp. 5939–5951, 2021.

[9] P. Hluska, M. Z. Li, T. Polishchuk, V. Polishchuk, and L. Sedov, "Strategic Demand Management in U-Space With RTTA Controls," in *Scitech*, 2025.

[10] D. Carramiñana, J. A. Besada, J. M. Molina, and A. M. Bernardos, "Scalable U-Space Authorization Workflow enabling Fairness and Capacity Management," in *14th SIDs*, Rome, Italy, 2024.

[11] Z. Liu, J. L. Munoz-Gamarra, and J. J. Ramos Gonzalez, "Characterization of strategic deconflicting service impact on very low-level airspace capacity." *Drones (2504-446X)*, vol. 8, no. 9, 2024.

[12] F. Causa and G. Fasano, "Towards a strategic deconfliction service based on multi-objective path planning for low altitude airspace," in *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC)*, 2024.

[13] J. L. Muñoz Gamarra, J. J. Ramos González, Z. Liu, and A. Sobejano Ventayol, "Impact of ussps performance in shared u-space volumes," *SESAR Innovation Days*, 2023.

[14] SESAR, "Concept of Operations for EuRopean UTM Systems," 2019.

[15] ——, "Concept of Operations for European U-space Services - Extension for Urban Air Mobility (CORUS-XUAM)," 2023.

[16] A. Martinez, P. Sanchez-Escalonilla, Y. Seprey, and V. M. Gordo, "DACUS Final Project Results Report," CRIDA, Deliverable D6.3, 2022.

[17] F. Ho, R. Geraldes, A. Goncalves, B. Rigault, A. Oosedo, M. Cavazza, and H. Prendinger, "Pre-flight conflict detection and resolution for uav integration in shared airspace," *IEEE Access*, 2019.

[18] Y. Tang, Y. Xu, R. Lv, and G. Inalhan, "Conflict probability based strategic conflict resolution for uas traffic management considering reasonable-time-to-act principle," *Transp Research Part C: Emerging Technologies*, 2025.

[19] Metropolis 2 consortium, "Metropolis 2: A unified approach to airspace design and separation management for U-space," Deliverable D3.1.

[20] L. Sedov, V. Polishchuk, and V. Bulusu, "Sampling-based capacity estimation for unmanned traffic management," in *36th DASC*, 2017.

[21] V. Bulusu, R. Sengupta, V. Polishchuk, and L. Sedov, "Cooperative and non-cooperative UAS traffic volumes," in *ICUAS'17*. IEEE, 2017.

[22] V. Bulusu and V. Polishchuk, "A threshold based airspace capacity estimation method for UAS traffic management," in *11th SysCon*, 2017.

[23] V. Duchamp, L. Sedov, and V. Polishchuk, "Density-adapting layers towards PBN for UTM," in *13th USA/Europe ATM R&D Seminar*, 2019.

[24] L. Sedov and V. Polishchuk, "Centralized and distributed UTM in layered airspace," in *8th ICRAT*, Barcelona, Spain, 2018, pp. 1–8.

[25] V. Bulusu, V. Polishchuk, and L. Sedov, "Noise estimation for future large-scale small uas operations," in *NOISE-CON*, 2017.

[26] M. Doole, J. Ellerbroek, and J. Hoekstra, "Estimation of traffic density from drone-based delivery in very low level urban airspace," *JATM*, 2020.

[27] SESAR 3 JU and EUROCONTROL, "European atm master plan (edition 2025)," https://www.sesarju.eu/masterplan, Dec. 2024, accessed 2025.

[28] V. Bulusu, R. Sengupta, and Z. Liu, "Unmanned aviation: To be free or not to be free?" in *7th ICRAT*, Philadelphia, PA, 2016.

[29] J. M. Hoekstra, "Designing for safety: The free flight air traffic management concept," Ph.D. dissertation, TU Delft, 2001.

[30] M. R. Jardin, "Analytical relationships between conflict counts and air-traffic density," *J. Guid. Control Dyn.*, vol. 28, no. 6, 2005.

[31] E. Sunil, J. Hoekstra, J. Ellerbroek, F. Bussink, D. Nieuwenhuisen, A. Vidosavljevic, and S. Kern, "Metropolis: Relating airspace structure and capacity for extreme traffic densities," in *ATMSeminar*, 2015.

[32] E. Sunil, J. Hoekstra, J. Ellerbroek, F. Bussink, A. Vidosavljevic, D. Delahaye, and D. Nieuwenhuisen, "How Do Layered Airspace Design Parameters Affect Airspace Capacity and Safety? The Influence of Traffic Structure on Airspace Capacity," in *7th ICRAT*, 2016.

[33] E. Sunil, J. Ellerbroek, J. Hoekstra, and J. Maas, "Modeling airspace stability and capacity for decentralized separation," *DEP*, vol. 3, 2017.

[34] M. Tra, E. Sunil, J. Ellerbroek, and J. Hoekstra, "Modeling the intrinsic safety of unstructured and layered airspace designs," in *ATMS*, 2017.

[35] V. Bulusu, V. Polishchuk, R. Sengupta, and L. Sedov, "Capacity estimation for low altitude airspace," in *17th AIAA ATIO*, 2017, p. 4266.

[36] "Icarus conclusions and results," 2022.

[37] L. Sedov, V. Polishchuk, and V. Acuna, "Altitude zoning for utm," in *10th SESAR Innovation Days*, 2020.

[38] A. M. Veytia, C. A. Badea, J. Ellerbroek, J. Hoekstra, N. Patrinopoulou, I. Daramouskas, V. Lappas, V. Kostopoulos, V. De Vries, J. van Ham *et al.*, "Metropolis II: Benefits of centralised separation management in high-density urban airspace," in *12th SIDs*, Budapest, Hungary, 2022.

[39] N. Patrinopoulou, I. Daramouskas, V. Lappas, V. Kostopoulos, A. M. Veytia, C. A. Badea, J. Ellerbroek, J. Hoekstra, V. de Vries, J. van Ham *et al.*, "Metropolis II: Investigating the future shape of air traffic control in highly dense urban airspace," in *30th MED*. IEEE, 2022.

[40] M. McPartland, G. Enea, T. Bonin, B. Dupree, C. Engholm, J. D. Cohen, A. Giulinao, N. Pena Ortiz, L. Garcia Insa, and S. Kim, "Towards the development of advanced air mobility high-resolution weather requirements," in *2026 AIAA Aviation Forum*. AIAA, 2026.

[41] A. M. Law, *Simulation Modeling and Analysis*, 5th ed. New York: McGraw-Hill Education, 2014.

[42] A. Evans, S. J. Wall, S. H. Young, and S. Pandey, "Simulating safety and efficiency impacts of airspace constraints in u-space airspace," in *42nd DASC*. IEEE, 2023.

[43] B. E. Barmore *et al.*, "Analysis of strategic conflict management approaches as applied to uam traffic," NASA Ames Research Center, Tech. Rep., 2023.

[44] T. J. DiCiccio and B. Efron, "Bootstrap confidence intervals," *Statistical Science*, vol. 11, no. 3, pp. 189–228, 1996.

[45] J. M. Flegal and L. Gong, "Relative fixed-width stopping rules for markov chain monte carlo simulations," *arXiv preprint arXiv:1303.0238*, 2013.

[46] P. Welch, "On the problem of the initial transient in steady-state simulation," IBM Watson Research Center, Tech. Rep., 1981.

[47] D. Raposo, D. Carramiñana, J. Besada, and A. Bernardos, "A Realistic UAS Traffic Generation Tool to Evaluate and Optimize U-Space Airspace Capacity," in *18th SOCO*. Springer, 2023, pp. 24–33.