This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Reduction of Air Traffic Complexity Using Trajectory-Based Operations and Validation of Novel Complexity Indicators

Tomislav Radišić, Doris Novak, and Biljana Juričić

***Abstract*— Airspace capacity is limited primarily by the saturation of air traffic controller's capacity, whose workload increases as air traffic complexity increases. Workload can be reduced through task automation by advanced controller tools. Automation and the development of novel controller tools is therefore one of the key aspects of future concepts of operations in European and American air traffic management systems. Implementation of trajectory-based operations (TBOs) has been proposed as a way to reduce workload, but few studies have examined how TBO affects air traffic complexity. This paper compares air traffic complexity experienced by ten air traffic controllers in a real-time simulation environment involving conventional operations and TBO. Analysis of subjective complexity scores collected in real time showed that TBO significantly reduced complexity when at least 70% of aircraft were flying according to TBO and when the airspace was occupied simultaneously by more than 15 aircraft. Subjective complexity scores were tested for correlation with 20 commonly used complexity indicators, and six indicators were used to generate a predictive linear model that performed well in conventional operations but less well under TBO. Therefore, we defined and experimentally validated two of seven novel TBO-specific complexity indicators. A second correlation model combining these two novel indicators with four already in use generated much better predictions of complexity than the first model.**

***Index Terms*— Air traffic complexity, trajectory-based operations, subjective complexity, complexity indicators.**

I. INTRODUCTION

GROWING demands on airspace are pushing the limits of current operational capacity. To meet increased demand, the Single European Sky ATM Research (SESAR) project recommends an increase in airspace capacity without an associated increase in ATM-related incidents or accidents, as well as decrease of environmental impact per flight and decrease in costs [1]. Achieving these goals will depend on adopting 4D trajectory management paradigm, which is the basis of the future ATM concept of operations called trajectory-based operations (TBO) [2].

In TBO, aircraft trajectories are agreed upon among airspace users, air navigation service providers (ANSPs), and airports. Airspace users are then obliged to fly their aircraft along the agreed trajectory with the required precision and accuracy in the four dimensions. ANSPs and airports, for their part, are obliged to facilitate that trajectory [2]. More precise and accurate ATM practices are deployed during flight planning and execution, which allows traffic conflicts to be solved on a strategic level rather than relying on air traffic controllers to solve them tactically. This approach improves management of human resources and infrastructure, and airspace and airport capacity; it also reduces financial and environmental costs of air traffic.

By reducing the number of conflicts that must be solved tactically, TBO is expected to reduce air traffic complexity, as stipulated in the SESAR WP 4 – En route operations [3]:

The goal of the SESAR concept is to deploy tools to manage complex situations in order to reduce complexity by strategic deconfliction measures within the new ATM system.

In the context of air traffic control, complexity was rarely clearly defined, perhaps due to assumed common knowledge. One notable exception is Meckiff (et al.) who stated that the air traffic complexity can be most easily defined as difficulty of monitoring and managing a specific air traffic situation [4]. Complexity is not the same as traffic density. Obviously, the number of aircraft in a sector (also known as density, traffic load, or traffic count) directly influences the air traffic complexity. This number, however, is not the only indicator of the level of complexity, especially if one wishes to compare different sectors of airspace [5]–[7].

Complexity is not a synonym for workload, although it has been proven multiple times that the increase in complexity results in the increase in workload which in turn limits the airspace sector capacity [8], [9]. Mogford *et al.* [6] reviewed numerous research articles in search of complexity and workload relationship. They concluded that complexity is actually the crucial factor for measuring controller workload. However, complexity and workload are not directly linked. Their relationship is mediated by several other factors, such as equipment quality, individual differences, and controller cognitive strategies, Figure 1.

Previously quoted expectation, stated in SESAR WP4 [3], that the strategic deconfliction measures will decrease complexity, is supported by several studies describing the interactions among air traffic complexity, air traffic controller workload and the resulting airspace capacity [6], [9], [10].

However, whether and how TBO affects air traffic complexity has yet to be tested directly through empirical studies. The objective of the present study was to examine directly, using

Manuscript received October 24, 2015; revised August 18, 2016 and January 26, 2017; accepted February 2, 2017. The Associate Editor for this paper was P. Ioannou. *(Corresponding author: Tomislav Radišić.)*

The authors are with the Faculty of Transport and Traffic Sciences, Department of Aeronautics, University of Zagreb, 10000 Zagreb, Croatia (e-mail: tradisic@fpz.hr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2017.2666087

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

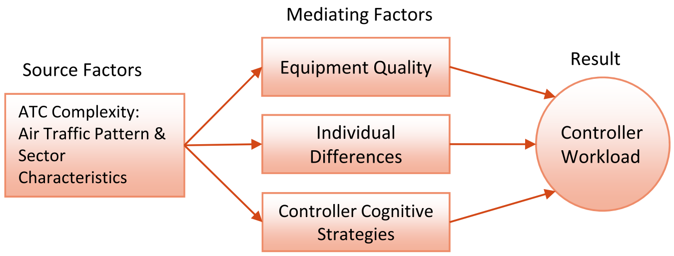


Fig. 1. The Relationship between ATC Complexity and Workload [6].

human-in-the-loop simulations with air traffic controllers, whether TBO leads to a reduction in the subjective air traffic complexity of en-route airspace sectors. Since this question is unexplored in the literature, the aim of this research is to use the subjective complexity scores to test the validity of commonly used objective air traffic complexity indicators for a TBO environment.

II. Methodology

To ensure more realistic conditions for comparing TBO and conventional operations, we opted for human-in-the-loop simulations over fast-time simulations or observational studies. Human-in-the-loop simulations have been used successfully to assess air traffic complexity [11], [12]. The simulations were carried out in a custom-built real-time air traffic control environment in the Department of Aeronautics of the Faculty of Transport and Traffic Sciences at the University of Zagreb. The experiments contained three scenarios which were based on the actual flight data: a scenario involving only aircraft flying by purely conventional operations, a scenario in which 30% of aircraft flew by TBO, and a third scenario in which 70% of aircraft flew by TBO. In addition, each scenario was created with one of three possible air traffic levels, giving a total of nine scenarios altogether.

A. Apparatus

A simulator used in this research was built and validated at the Laboratory for Control of Air Navigation at the Department of Aeronautics, Faculty of Transport and Traffic Sciences, University of Zagreb. The development of the research simulator started after the review of the commercial off-the-shelf simulators showed that it was impossible to perform this kind of research on the existing equipment. The main issues with the existing simulators were inability to simulate 4D trajectories, difficulty in measuring and storing all of the necessary data, and costly customization. Also, it was concluded that a custom-built simulator could later be adapted and reused for future research.

The ATC simulator used in this research has the following characteristics:

- **Accurate and versatile aircraft models**. EUROCONTROL's Base of Aircraft Data (BADA) Aircraft Performance Model (APM) was chosen as a starting point for aircraft model. Its main advantages are support for many different aircraft types, easy implementation, and excellent documentation. Base of Aircraft Data (BADA) is a database of aircraft data developed and updated by EUROCONTROL Experimental Centre (EEC). As mentioned by Nuić *et al.* [14] the aircraft performance information provided in BADA *'is designed for use in trajectory simulation and prediction in ATM research as well as for modeling and strategic planning in ground ATM operations'*. It provides ASCII files containing operation performance parameters for 405 aircraft types – out of which 150 are original aircraft types and 255 are equivalent aircraft types. BADA, however, provides only aircraft performance modelling so the models of aircraft dynamics and FMS had to be developed from the start.

Fig. 2. ATC Simulator Working Environment.

- **Realistic working environment**. It had to be similar to the realistic working environment to which the ATCOs are used to. This includes the layout of the radar screen, auxiliary screens, a keyboard, a mouse, and some communication switches. The user interface had to be similar to the existing ATC simulators and workstations to give the ATCOs a smooth transfer to the simulator (without an extensive training) (Fig. 2.).
- **Representative ATC tool operation.** The simulator supports the following tools: strip-less flight progress monitoring system (for 3D and 4D navigation), map display configuration tool, range and bearing lines, level and SSR code filters, separation (SEP) tool, area proximity warning (APW), short-term conflict alert (STCA), separation infringement alert, display tools, flight profile tool etc.

Before the simulator was used in the research, it was validated using a series of tests. The user interface and tool operation were validated by comparison with operational ATC systems and by expert assessment, the aircraft models were validated by comparison with real-life flight data collected from quick access recorders, and working environment was validated by the licensed ATCs [13].

B. Airspace

The airspace used for simulation experiments was the Croatia Upper North sector (Fig. 3, dashed light blue line). This airspace was chosen because the air traffic controllers participating in the simulations were quite familiar with it after working for several years for the Croatian ANSP (Croatia Control Ltd). This familiarity helped ensure that participants could accurately assess differences in air traffic complexity, it

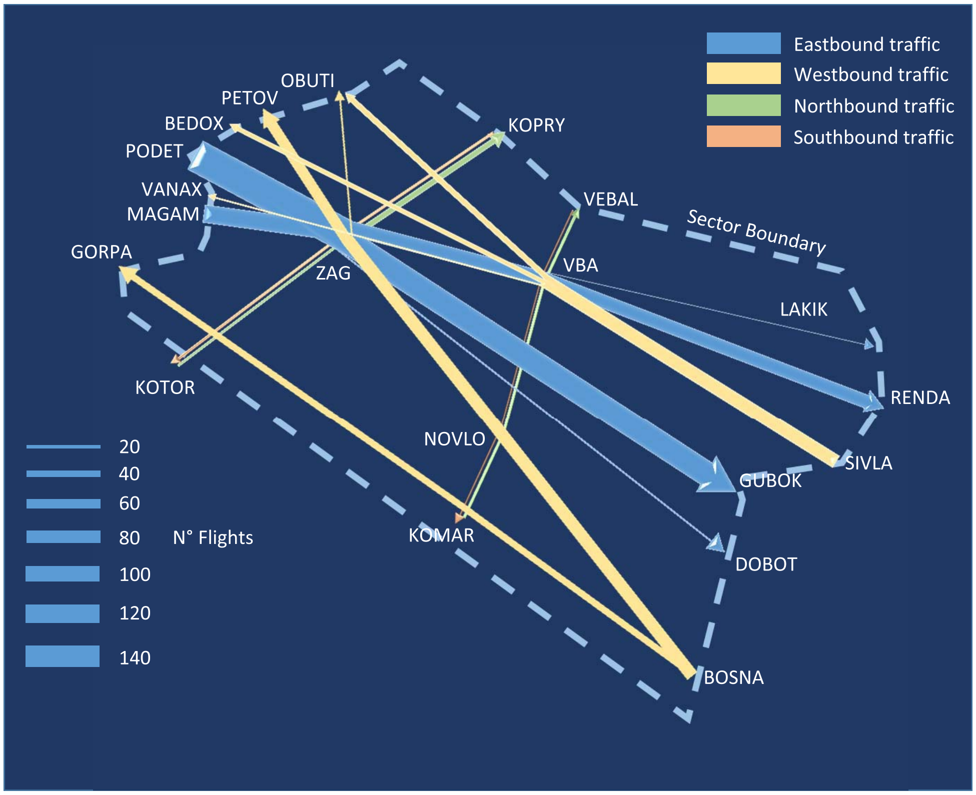


Fig. 3. Most frequently used routes in Croatia Upper North sector.

reduced the need for pre-simulation training and it eliminated the possible confounder of different learning rates.

The Croatia Upper North airspace sector is class C airspace dominated by the ZAG VOR/DME, through which most of the RNAV routes in the sector pass.

During the simulations, participants were required to adhere to the Flight Level Allocations and Special Procedures (FLAS), coordination points and transfer of control points stipulated within the Letters of Agreement governing the transfer of traffic between surrounding area control centers (ACC) and Zagreb ACC. These conditions are intended to ensure that flights crossing a boundary between ACCs can land properly at the desired airport or join seamlessly with existing traffic flows.

C. Traffic

To ensure that air traffic in the simulation experiments was as realistic as possible for the selected airspace, historic traffic data were obtained from EUROCONTROL for a single summer day (30 August 2013), which was selected because traffic varied substantially throughout that day. Of the 661 flights through the Croatia Upper North airspace sector, approximately 70% involved commercial medium jets and 10% involved heavy jets. The remaining flights involved primarily regional turboprops and business jets.

The routes used most frequently on the selected day connected the southeast and the northwest of Europe (Fig. 3). Overall, 90% of flights moved on a southeast-northwest axis, and the remaining 10% on a northeast-southwest axis. More than 50% of all flights followed one of the five most frequently used routes. Other European airspace sectors have different configurations of the traffic flows and therefore some results of this research will not be directly applicable (e.g. values of regression coefficients calculated for evaluation of objective complexity indicators) but other results (e.g. subjective air traffic complexity scores and list of validated objective complexity indicators) are very likely to be applicable to some sectors with different traffic flows.

Nine different scenarios were conducted during the experiment, involving three operations environments (conventional, 30% TBO, or 70% TBO) and three air traffic levels (low, high, or future). Traffic data were sampled during off-peak periods to build scenarios with low traffic levels, and from peak periods to build scenarios with high traffic levels. In the scenarios featuring a future traffic level, additional flights were added to routine traffic to give rise to an unrealistically high aircraft count. In addition, the proportion of aircraft climbing or descending was higher than in the scenarios with low or high traffic levels. The aim of the future scenarios was to expose controllers to complexity beyond what can be expected nowadays and beyond what the controllers had previously experienced in their careers.

The additional flights added to the routine, real data-based flights in the future traffic scenarios were generated in a semi-stochastic manner. Firstly, a route was chosen randomly such that the probability of selecting a given route was equal to the frequency with which it was flown on the selected day. Secondly, aircraft type was randomly chosen so that the probability of selecting a given type reflected the actual aircraft distribution on that day. Thirdly, an appropriate flight level was chosen for the selected route based on semi-circular flight level allocation rules (eastbound flights fly at odd flight levels, whereas westbound flights use even flight levels).

Finally, the time of entry into the sector for the given flight needed to be generated in a way to ensure that no conflicts occurred within 5 minutes of the aircraft's entry into the sector. This was mandated by Letters of Agreement with the neighboring ACCs or by internal procedures if the aircraft entered from another sector within the ACC. This was achieved by randomly generating the time of entry which satisfied this condition for that particular aircraft and compliance was checked by running fast-time simulations. This method ensured that the artificially generated flights showed approximately the same pattern as the real ones, thereby minimizing unrealistic traffic flows.

In the scenarios in which 30% or 70% of aircraft flew by TBO, aircraft to be converted to TBO were selected so that the proportion of TBO aircraft would remain nearly constant throughout the simulation. Some fluctuation was unavoidable, since conversion of a single flight to TBO could change the relative proportions of conventional and TBO aircraft by up to 10%. Post-hoc analysis showed that the relative proportions remained constant within $\pm 15\%$ during 95% of the low simulation scenario, excluding the very beginning and end of the simulations, and that they remained constant within $\pm 5\%$ during 95% of the other scenarios. TBO aircraft were then deconflicted among themselves in order to simulate strategic deconfliction, one of the main features of TBO [3]. Deconfliction was performed in fast-time simulations by adjusting times of sector entry. If deconfliction could not be achieved by modifying entry times by 30 seconds or less, then the conflict was solved by changing the flight level for one of the aircraft. If neither aircraft could change level, e.g. because of other traffic or performance limitations, then trajectories were adjusted slightly by inserting new waypoints into the flight plans (which amounts to vectoring). After strategic deconfliction, controllers honored the agreed business trajectory of TBO aircraft without making any adjustments to it at a tactical level. The aircraft flying according to TBO were handed over in the same way as conventional aircraft, i.e. via voice communication.

D. Participants

All 10 controllers (8 male, 2 female; mean age, 31; age range, 27-34) who participated in the simulations were recruited from the Croatian ANSP, Croatia Control Ltd. These controllers had an average of 7 years' experience (range, 4-11) working at the Croatian ANSP, and an average of 5 years (range, 2-9) had passed since they had received their air traffic control license. They all had an extensive experience of controlling the traffic in the Croatia Upper North airspace sector. Two additional participants were aeronautical engineers with master's degrees who had completed formal training as air traffic controllers and who assisted in the development and testing of the experiment scenarios. These two participants were not involved in the experiments in which subjective air traffic complexity was measured.

Before the execution of the experiment, each controller received a brief training in order to become accustomed with the simulator interface and operational procedures. The training consisted of an introductory lecture, pre-simulator briefing, trial simulator runs, and a post-simulator briefing. The introductory lecture covered basic topics in air traffic complexity, the subjective complexity rating scale used in the study, TBO, simulator tools and features, airspace, experiment scenarios, and operational procedures. The trial simulator runs lasted at least 90 min and involved two scenarios, one with conventional operations and one with TBO. All participants declined to participate in additional training simulations that were offered, indicating that they felt sufficiently comfortable with the simulator operations.

One of the authors participated as the pseudo-pilot in all experiments. The controller could communicate with the pseudo-pilot only via a headset. Since the pseudo-pilot had to take on the role of air traffic controllers in other air traffic control units, to facilitate coordination an assistant to the pseudo-pilot participated in the scenarios involving high and future traffic levels; this assistant was one of the aeronautical engineers with a master's degree. The experiments did not involve planner controllers, only the executive ones.

In order to prevent learning or other unwanted effects due to the order of scenarios, each controller was exposed to the three operations environments (conventional, 30% TBO, 70% TBO) in random order. Within each type of operations environment, however, the three traffic loads (low, high, future) were always presented in the same order. This was intended to help controllers assess complexity more consistently.

The scenarios began with an empty airspace to avoid possible confounding due to the absence of a hand-over step. Normally, controllers starting their shift with an occupied airspace would have the benefit of observing traffic for 15 minutes or so while another controller manages it. Since such a hand-over was not a part of our design, we did not want to start our scenarios with an occupied airspace in order to avoid inducing initial disorientation and increased workload that might affect baseline complexity scoring.

E. Air Traffic Complexity Rating

The controllers were asked to subjectively rate air traffic complexity throughout the simulation, using a modified ATWIT scale [15] that we named the Air Traffic Complexity Input Technique (ATCIT) [16]. The ATCIT scale features seven levels of complexity (Table I).

The levels of subjective complexity on this scale reflect primarily the controller's self-assessment of situational awareness, while also taking into account aircraft-aircraft and aircraft-airspace interactions. Before using this scale, the controllers were briefed about the objectives of the ATCIT scale and the meaning of 'complexity', 'interaction', and 'situational awareness'.

During each simulation run, a Subjective Complexity Measurement (SCM) tool opened every 2 minutes, accompanied by non-intrusive aural notification. The tool presented

TABLE I
ATCIT SCALE [14]

Complexity Score	Description
1	No complexity – no traffic
2	Very low complexity – very little traffic, no interactions
3	Low complexity – situation and interactions obvious at a glance
4	Somewhat low complexity – firm grasp of the situation, interactions are anticipated and prepared for
5	Somewhat high complexity – aware of the situation, interactions are handled in time
6	High complexity – having trouble staying aware of all interactions, occasionally surprised by unnoticed interactions and conflict alerts
7	Very high complexity – losing situational awareness, unable to track all interactions, responding reactively

7 buttons labelled 1-7, and the controller had to click on the button most closely matching the perceived level of air traffic complexity. For reference, the scale was permanently displayed on a piece of paper beside the radar screen. Each assessment was time-stamped and automatically recorded. Throughout the simulation run, at one-second intervals, the data on objective complexity indicators were calculated, time-stamped, and recorded. The purpose of this data was comparison with subjective complexity indicators.

In order to identify commonly used indicators of air traffic complexity that we might validate for the use with TBO, we conducted a literature review that highlighted more than 100 factors (gathered in two comprehensive reviews in [17] and [18]). Since our focus was on air traffic complexity under ideal conditions, we disregarded indicators related to weather. All simulation runs were performed in the same airspace sector and no attempts were made to compare it with other sectors, therefore indicators related to the airspace complexity were filtered out as well. Performing the whole experiment in the same sector probably made some of the results not applicable generally (e.g. values of regression coefficients calculated for evaluation of objective complexity indicators) but other results (e.g. subjective air traffic complexity scores and list of validated objective complexity indicators) were very likely applicable to other sectors because airspace-specific data was not used in analysis. Previous research by Kopardekar showed that this method produced models which performed well in other sectors [19].

We also disregarded indicators related to emergencies, government or military aircraft, and equipment malfunctions. Since our focus was on en-route operations, we further discarded factors related to terminal operations, departure/arrival traffic flows, approach procedures, and airports. In the end, we developed a list of clearly defined, previously experimentally validated objective complexity indicators:

- Aircraft count;
- Volume of convex hull (described by aircraft position);
- Aircraft density I (based on sector volume);
- Aircraft density II (based on convex hull volume);
- Aircraft density II squared (based on convex hull volume and the squared number of aircraft);
- Separation criticality index;
- Number of aircraft with horizontal separation less than 8 NM;
- Inverse of minimum horizontal separation in the same vertical neighborhood;
- Inverse of minimum vertical separation in the same horizontal neighborhood;
- Ratio of standard deviation to mean value for ground speed;
- Ratio of mean aircraft distance to number of aircraft;
- The intersection angle for aircraft less than 13 NM apart;
- Fraction of aircraft with fewer than 600 seconds to conflict;
- Fraction of aircraft climbing;
- Fraction of aircraft descending;
- Fraction of aircraft either climbing or descending;
- Number of aircraft pairs at a 3D Euclidean distance less than 5 NM;
- Number of aircraft pairs at a 3D Euclidean distance of 10-15 NM;
- Variation in aircraft headings relative to the sector axis; and
- Standard deviation of aircraft headings.

In addition to drawing on these literature-based complexity indicators, we wanted to be in a position to identify new, TBO-specific complexity indicators based on data from the simulation experiments (see section III.C). Therefore data on the state of all aircraft were recorded at 1-second time step in the simulation. These state data included, but were not limited to, position, velocity, heading, mass, pitch, bank, throttle, drag, climb mode, acceleration mode, assigned flight level, speed, heading, and route.

F. Study Limitations

To ensure that the most rigorous analyses could be performed within the constraints of this study, only en-route operations were considered, since simulating terminal airspace conditions would require a completely different set of simulation scenarios and participants. In addition, only nominal operations were considered, excluding inclement weather, emergencies, and special operations (military, government, medical). Our intention was not to model unique, off-nominal situations but rather situations reflective of routine ATM.

The controllers in our simulations were required to honor the business trajectory contracts. The conflicts among aircraft flying according to conventional operations were solved tactically, the conflicts among TBO aircraft were solved strategically, and the conflicts between conventional and TBO aircraft were solved tactically by honoring the business trajectory contract of the TBO aircraft, meaning that the route of the conventional aircraft was always adjusted. We adopted this

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

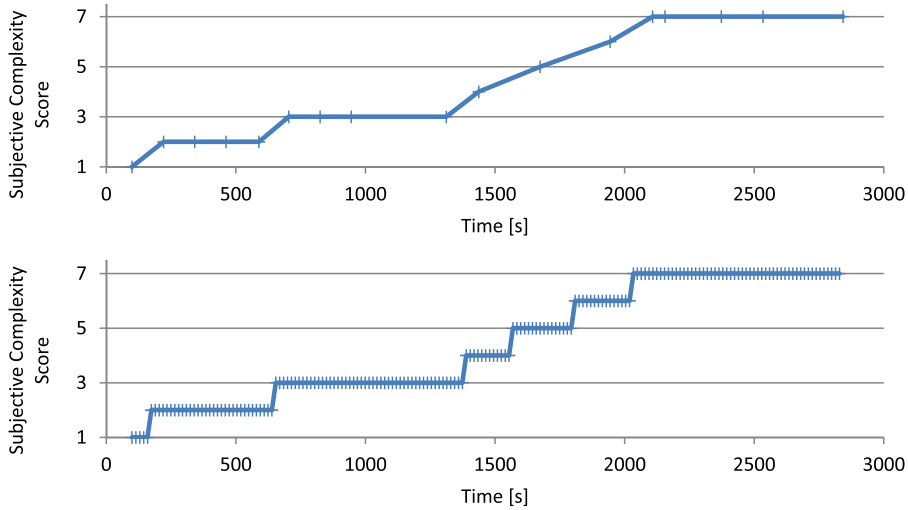


Fig. 4. Example of subjective complexity scores before (top) and after (bottom) resampling.

rule as a simplifying measure to ensure consistency in our data from different controllers. In practice, it seems more likely that controllers would have the freedom to choose how they solve conflicts in mixed operations. Nevertheless, although SESAR documents do not currently recommend particular conflict resolution procedures in mixed operations, it is possible that a procedure similar to our rule may ultimately be implemented to encourage TBO adoption among aircraft operators.

The controllers had available only a limited set of tools, which was intentional to prevent equipment complexity from influencing perceptions of air traffic complexity. The limited set of tools nevertheless seemed adequate because some of the tools were never used by the controllers during the experiment, and no controller expressed a desire for additional tools.

III. Results

A. Data Recording and Processing

Throughout the experiment, three types of data were recorded: raw aircraft state data (every 1 second), which we planned to use to develop potentially novel, TBO-specific complexity indicators; the data on 20 complexity indicators (calculated every 1 second) that we identified from the literature (section II.E); and subjective complexity scores (every 2 minutes). Each participant was asked to perform nine simulation scenarios, each lasting approximately 50 minutes, implying 25 complexity scores per participant per simulation or 2250 complexity scores across all 10 participants. Of these, only 1997 complexity scores were actually obtained because one participant had to withdraw from the study for personal reasons after completing only 7 simulation runs (accounting for 50 lost scores), and some participants did not enter complexity scores as soon as they were prompted, sometimes due to intense focus on controlling the traffic (accounting for 146 lost scores). In addition, 6 participants did not complete all the simulation runs with a future traffic level because they lost situational awareness, leading to separation minima infringement (accounting for 57 lost scores). These missing 57 data points were assigned the maximum ATCIT score of 7 and incorporated into our final analysis.

We used the nearest-neighbor interpolation to fill in the 146 gaps in subjective complexity caused by late input of complexity scores. Re-sampling the scores at 15-second intervals increased the number of samples for those parts of the simulator run when scoring gaps were more prominent. Since most scoring gaps occurred during the parts of the simulation scenarios involving high workload, the interpolation procedure slightly increased mean complexity scores.

Fig. 4 shows an extreme example of the changes caused by re-sampling. Each short vertical line indicates a sampling point. The upper panel shows that sampling was not uniform throughout the simulation. In the resampled lower panel, the sampling is distributed evenly.

For the particular data shown in Fig. 4, resampling increased mean subjective complexity from 4.17 to 4.43 but reduced the standard deviation from 2.21 to 2.03.

Since the simulations began with no aircraft in the airspace, the controllers had some time until the aircraft count reached the relevant traffic level, defined as >10 aircraft in low-traffic scenarios, >15 aircraft in high-traffic scenarios and >20 aircraft in future-traffic scenarios (Fig. 5).

The cut-off times selected for extracting complexity scores (shaded region in Fig. 5) were determined for each simulation scenario in order to capture the features differentiating the 3 scenarios from each other. Since low-traffic scenarios and high-traffic scenarios were created from historic traffic data (off-peak and peak traffic, respectively), the cut-off was determined as the time at which the aircraft count in two scenarios naturally diverged. For future-traffic scenarios, the cut-off was determined as the time at which the aircraft count increased beyond the historic peak levels. This procedure filtered out the data which was similar to the scenario with lower aircraft count and which was, therefore, not relevant for the given scenario with higher aircraft counts.

Therefore, subjective complexity scores obtained at the beginning of each simulation were omitted from the analysis. For the same reason, the scores obtained at the end of

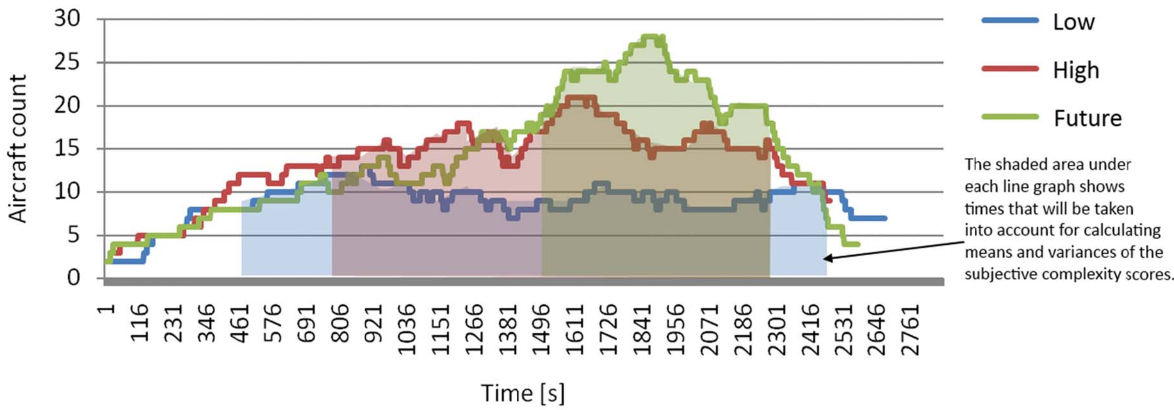


Fig. 5. Example of filtering subjective complexity scores.

each simulation were also discarded as aircraft counts began converging for all 3 scenarios.

The number of subjective complexity scores retained after applying these cut-offs decreased as aircraft count increased (Fig. 5).

B. Hypothesis Testing

Our hypothesis in these experiments was that TBO would lead to lower air traffic complexity than conventional operations in en-route airspace sectors. This hypothesis can be expressed mathematically as in Eq. 1, as hypothesis of equal means (μ), while alternative one is hypothesis that means are different:

$$H_0 : \mu_c = \mu_{TBO30\%} = \mu_{TBO70\%} \tag{1}$$

$$H_A : \mu_c = \mu_{TBO30\%} \vee \mu_c > \mu_{TBO70\%}. \tag{2}$$

The hypothesis was tested in three stages: first, means were compared between conventional and TBO scenarios in simulations with low traffic level; next, this process was repeated for simulations with high and future traffic levels. Table II shows mean subjective complexity scores for each participant and each scenario.

The hypothesis was tested using one-way repeated-measures ANOVA independently for each of the three traffic levels. Confidence intervals were adjusted using Bonferroni's method; if the result was non-significant, the least significant difference was also calculated [20], [21].

Applying Mauchly's [22] test to scores from low-traffic scenarios showed that the assumption of sphericity was violated [$\chi 2(2)$ = 14.116, p = 0.001], so the degrees of freedom were corrected using the Greenhouse-Geisser [23] estimate of sphericity ($\varepsilon = 0.547$). The results showed no significant effect of TBO on subjective complexity scores [F(1.094, 9.843) = 0.980, p = 0.355].

Mauchly's test for data from high-traffic scenarios indicated that the assumption of sphericity was valid [$\chi 2(2)$ = 0.378, p = 0.828], so the degrees of freedom were not corrected. The results showed that TBO was associated with significantly lower subjective air traffic complexity scores [F(2, 18) = 14.707, p < 0.001]. Post-hoc analysis showed that the mean difference was significant only between 0% TBO and 70% TBO (MD = 0.79, p = 0.001), and between 30% TBO and 70% TBO (MD = 0.625, p = 0.007).

Since subjective complexity was assessed on an ordinal scale, we confirmed our results with high-traffic scenarios using the non-parametric Friedman test [24]. TBO significantly reduced subjective complexity at a high traffic level [$\chi 2(2)$ = 15.8, p < 0.001], and post-hoc analysis using the Wilcoxon signed ranks test [25] indicated a significant difference between 0% TBO and 70% TBO (Z = 2.803, p = 0.005), and between 30% TBO and 70% TBO (Z = 2.805, p = 0.005), but not between 0% TBO and 30% TBO (Z = 1.580, p = 0.114).

TABLE II

MEANS AND VARIANCES OF SUBJECTIVE COMPLEXITY SCORES

Participant	Stat.	Low traffic			High traffic			Future traffic		
		0% TBO	30% TBO	70% TBO	0% TBO	30% TBO	70% TBO	0% TBO	30% TBO	70% TBO
1	μ	2.69	2.88	2.87	3.46	3.50	2.97	5.85	-'	-'
	σ²	0.34	0.10	0.11	0.25	0.25	0.43	0.41	-'	-'
2	μ	2.05	2.00	1.99	2.97	3.90	2.61	6.26*	5.02	3.60
	σ²	0.05	0.00	0.01	0.69	1.10	0.37	0.88	1.02	0.28
3	μ	3.31	1.83	1.88	4.24	4.11	2.97	6.14*	4.63	5.83*
	σ²	0.22	0.14	0.10	0.79	0.56	0.03	1.21	0.24	3.11
4	μ	1.94	1.93	1.77	2.74	2.53	2.12	6.48*	4.75	3.71
	σ²	0.17	0.07	0.18	0.48	0.25	0.11	1.03	0.94	1.80
5	μ	3.19	3.38	3.35	4.32	4.17	4.03	6.09	6.29	5.98
	σ²	0.56	0.58	0.58	0.73	0.87	0.10	0.65	0.80	0.52
6	μ	2.75	2.71	2.62	4.50	3.80	3.07	6.83*	6.09	5.83*
	σ²	0.19	0.33	0.24	0.32	0.75	0.06	0.17	0.68	1.86
7	μ	2.16	2.00	2.27	4.04	2.90	2.77	6.34	6.48	6.05
	σ²	0.13	0.00	0.20	0.63	0.09	0.18	0.66	0.41	0.76
8	μ	2.05	2.00	2.00	3.00	2.74	2.47	4.20	4.51	4.74
	σ²	0.05	0.00	0.00	0.53	0.20	0.25	0.51	0.25	0.23
9	μ	1.15	1.00	1.00	2.48	2.58	1.26	6.98*	6.14*	5.31*
	σ²	0.13	0.00	0.00	1.28	2.39	0.20	0.02	1.40	1.12
10	μ	2.19	2.31	2.16	3.36	3.20	2.93	6.08*	5.62	4.74
	σ²	0.29	0.22	0.13	0.88	0.28	0.06	1.17	0.80	0.23

μ – mean
σ² – variance
' – did not attempt the scenario
* – did not complete the scenario

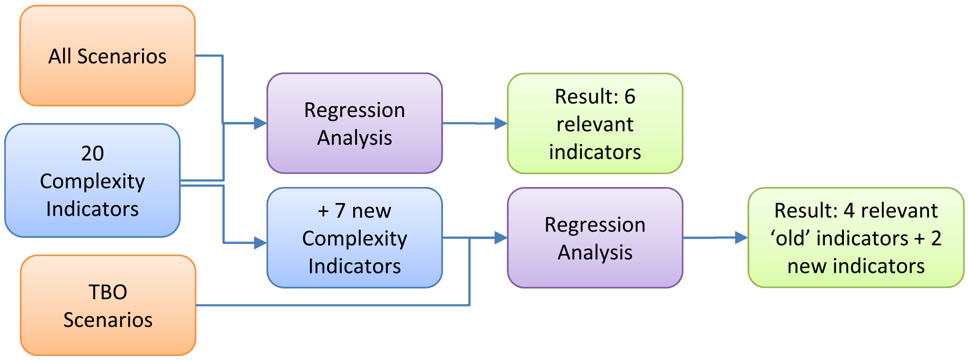


Fig. 6. Overview of regression analysis procedure.

Applying Mauchly's test to scores from future-traffic scenarios indicated that the assumption of sphericity was valid [$\chi 2(2) = 2.948$, $p = 0.229$]. Analysis showed that TBO significantly reduced subjective air traffic complexity scores [$F(2, 16) = 6.280$, $p = 0.01$]. However, post-hoc analysis with Bonferroni adjustment for multiple comparisons showed no significant difference between pairs of scenarios: 0% TBO vs 70% TBO (MD = 1.070, $p = 0.071$); 0% TBO vs 30% TBO (MD = 0.655, $p = 0.096$); and 30% TBO vs 70% TBO (MD = 0.415, $p = 0.444$). On the other hand, post-hoc analysis using the less stringent least significant difference to adjust for multiple comparisons showed significant differences between 0% TBO and 70% TBO (MD = 1.070, $p = 0.024$) and between 0% TBO and 30% TBO (MD = 0.655, $p = 0.032$), but not between 30% TBO and 70% TBO (MD = 0.415, $p = 0.148$).

Non-parametric Friedman testing showed a significant effect of TBO on subjective complexity at future traffic levels [$\chi 2(2) = 6.889$, $p = 0.032$], while post-hoc analysis using the Wilcoxon signed ranks test indicated a significant difference between 0% TBO and 70% TBO ($Z = 2.192$, $p = 0.028$), but not between 30% TBO and 70% TBO ($Z = 1.599$, $p = 0.110$) or between 0% TBO and 30% TBO ($Z = 1.955$, $p = 0.051$).

These results suggest that TBO can significantly reduce subjective air traffic complexity, but only when the traffic level and proportion of TBO aircraft are high.

As a test of the robustness of our ACTIT scale and simulation procedure, we examined consistency of air traffic complexity scores between different controllers for the same scenarios and between different moments of the same scenario for the same controller. In both cases consistency was low. Different controllers often assigned substantially different scores to the scenarios with the same traffic level, and in rare cases some controllers assigned the same complexity score throughout an entire scenario; in one case, this led to the unlikely situation in which an airspace with only two aircraft received the same score as an airspace with 12 aircraft.

C. Evaluation of Objective Complexity Indicators

Given that the present study is one of the first to examine in detail how TBO may affect air traffic complexity, we wanted to examine whether commonly used objective complexity indicators are likely to be suitable for TBO. We used linear regression, since it has already proven useful for other researchers who tried modelling air traffic complexity using objective complexity indicators, e.g. in [12], [26], and [27]. First, 20 commonly used and validated indicators (section II.E) were tested against our subjective complexity data from all nine scenarios. Second, the regression was repeated – this time only on TBO simulations – using the same 20 indicators as well as 7 potentially new, TBO-specific indicators (Fig. 6). Regression was performed in a step-wise manner with stepping criteria set to $p < 0.05$ for inclusion and $p > 0.10$ for exclusion of the indicator.

In the first step, using only 20 commonly used complexity indicators, the model that best predicted subjective complexity scores among all simulations showed the following characteristics: $R = 0.746$; $R^2 = 0.556$; R^2-Adjusted $= 0.554$. The R^2-Adjusted is used here because it takes into account the number of explanatory terms in a model relative to the number of data points [28]. This best model contained the following indicators, listed in order of importance: number of aircraft, fraction of aircraft climbing or descending, heading variance, number of aircraft pairs at a 3D Euclidean distance less than 5 NM, number of aircraft near the sector boundary (<10 NM), and the ratio of mean aircraft distance to number of aircraft.

At low traffic levels, air traffic complexity scores often remained constant or nearly so throughout the scenario. Presumably the controllers perceived low complexity within a narrow dynamic range, leading them to give just one score (1 or 2) for the entire scenario. This threatened the robustness of our model, which we assessed by applying the model to each of the scenarios individually and to groups of scenarios at the same traffic level (Table III).

This sensitivity analysis showed that less complex simulations (all low-traffic scenarios + high-traffic 70% TBO scenarios) had much lower R^2 values than more complex scenarios. This likely reflects the uniform ATCIT scores given by participants in the low-complexity scenarios. The sensitivity analysis also revealed a more subtle effect: among scenarios at the same traffic level, R^2 values decreased slightly as the proportion of TBO aircraft increased (Fig. 7). This finding suggests that the selected complexity indicators show greater predictive power in conventional operations than in TBO, which led us to attempt to identify more suitable TBO-specific indicators.

We defined potential TBO-specific complexity indicators to capture the interaction between TBO and conventional aircraft,

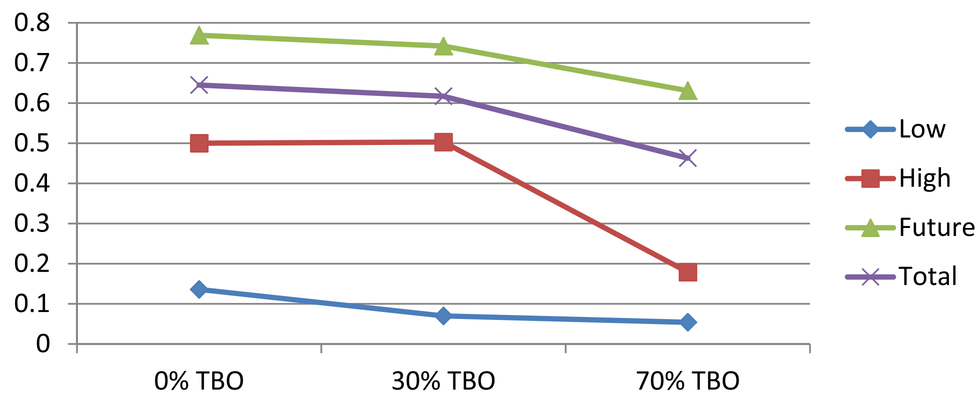


Fig. 7. Values of R^2–Adjusted for different scenarios.

TABLE III
RESULTS OF REGRESSION ANALYSES

Simulation Run		R	R^2	R^2-Adjusted	Std. Err. of the Estimate
Low	0% TBO	0.380	0.144	0.136	0.6492
	30% TBO	0.272	0.074	0.070	0.7638
	70% TBO	0.242	0.059	0.054	0.6901
	ALL	0.314	0.099	0.093	0.7016
High	0% TBO	0.709	0.503	0.500	0.7819
	30% TBO	0.711	0.506	0.503	0.7672
	70% TBO	0.428	0.183	0.179	0.7438
	ALL	0.616	0.380	0.378	0.8204
Future	0% TBO	0.880	0.775	0.769	0.8337
	30% TBO	0.864	0.746	0.742	0.7805
	70% TBO	0.799	0.639	0.631	0.9112
	ALL	0.833	0.694	0.690	0.8890
All	0% TBO	0.805	0.649	0.645	0.8101
	30% TBO	0.788	0.621	0.617	0.8036
	70% TBO	0.684	0.468	0.463	0.8298
ALL		0.746	0.556	0.554	0.8507

given that TBO aircraft were strategically deconflicted, so they did not interact with one another.

The following indicators were tested in linear regression against data from TBO simulations [9]: fraction of TBO aircraft, number of conflicts between conventional aircraft and TBO aircraft (during 600 seconds), fraction of TBO aircraft climbing or descending, fraction of conventional aircraft climbing or descending, number of conventional aircraft at a 3D Euclidean distance of less than 5 NM from TBO aircraft, number of conventional aircraft at a 3D Euclidean distance of 5-10 NM from TBO aircraft, and number of conventional aircraft at a 3D Euclidean distance of 10-20 NM from TBO aircraft.

Multiple step-wise linear regression analysis was performed again, but this time with all 27 indicators and with only those scenarios with TBO traffic. The resulting 6-factor model included four complexity indicators that we identified from the literature and two new TBO-specific indicators, listed in order of importance: number of aircraft, number of conflicts between conventional aircraft and aircraft flying according to TBO (aggregated over 600 seconds), fraction of aircraft in climb or descent, number of aircraft near sector boundary ($<$10 NM), fraction of TBO aircraft, and number of aircraft pairs at 3D Euclidean distance less than 5 NM.

The new model showed the following characteristics: $R = 0.833$; $R^2 = 0.693$; R^2-Adjusted $= 0.691$. This new model containing two novel complexity indicators correlated better with scores from TBO simulations than the original model comprising only indicators from the literature, which gave an R^2-Adjusted of 0.617 for 30% TBO and 0.463 for 70% TBO.

IV. DISCUSSION

Analysis of ATCIT scores in this human-in-the-loop simulation experiment suggests that the air traffic complexity perceived by controllers decreases as the proportion of TBO aircraft increases, but only at higher traffic levels. This can be attributed to the reduced number of aircraft-aircraft interactions. The aircraft flying according to TBO were strategically deconflicted, thus reducing the number of possible interactions among aircraft at tactical level. This effect was not noticeable to controllers at lower traffic levels, when few interactions were present.

It may be that with a larger number of controllers and scenarios, we would be able to detect an effect of TBO on subjective air traffic complexity under conditions with fewer aircraft and smaller proportions of TBO aircraft. Our study shows the validity of this approach and so justifies larger, more extensive investigations in the future.

Our results from the scenarios with future traffic levels should be regarded with caution, partly as a result of sample size issues, especially due to the withdrawal of one participant from the study. The calculation of the required sample size for scenarios with future traffic levels, using the same parameters (effect size, significance level) as those in the scenarios with high traffic levels, showed that the smallest sample size which could be used to detect the effect of TBO on subjective air traffic complexity was 11 controllers. Therefore, if the effect size increased from high to future traffic level scenarios (as it did from low to high traffic level scenarios) it might be that one missing controller in scenarios with future traffic levels was crucial for proving the research hypothesis (for future scenarios). However, though we detected no significant difference between pairs of scenarios in the scenarios with future traffic levels when we imposed the relatively strict

Bonferroni multiple-comparisons correction, we did detect an effect when we used the less strict least significant difference method. In this case, the results were similar to those for simulations at high traffic level. In addition, non-parametric testing gave different results from parametric testing.

One limitation to our approach is seen in the low consistency that we observed within and among the controllers. While this poor consistency did not affect the statistical analyses, they make it impossible to compare different controllers' perceptions of complexity, which may be valuable for more detailed studies in the future. It is unclear how such consistency can be improved using human-in-the-loop simulations. Participants could perhaps receive more extensive training, such as with static radar images of traffic situations, in order to help them calibrate their scores and encourage them to use the full range of the ATCIT scale. On the other hand, imposing an explicit calibration system may introduce other biases that cause controllers to behave non-naturally in the simulations, threatening external validity.

The purpose of the second part of our study was to determine whether current objective complexity indicators are suitable for assessing complexity in TBO. Regression analysis showed that a model with six objective complexity indicators can explain the variance in subjective complexity scores with an R^2-Adjusted of 0.556. This value can be increased to 0.75 by selecting scenarios showing greater variance in subjective complexity scores, thereby compensating to some extent for poor rater consistency. These results are comparable to the unadjusted R^2 of 0.69 (probably smaller if adjusted) reported by Kopardekar et al. in their modelling of air traffic complexity in conventional operations [19]. Our finding that R^2 reduced with increasing proportion of TBO aircraft, regardless of traffic level, suggests that currently used complexity indicators, while providing reasonable predictions, are not suitable for TBO.

As a first step towards identifying and validating TBO-specific objective air traffic complexity indicators, we defined 7 candidate complexity indicators designed to explain the interaction between conventional and TBO aircraft. Multiple linear regression with all 27 indicators (20 old and 7 new) gave a model with 6 indicators (4 old and 2 new) with an R^2-Adjusted of 0.691. Thus, these 2 TBO-specific indicators significantly improved prediction of subjective complexity. It is likely that larger studies in the future will help refine these TBO-specific indicators and identify additional ones. Such studies should perhaps include non-linear modeling that includes objective complexity indicators in order to increase predictive power.

V. Conclusion

These human-in-the-loop experiments, which we believe to be the first direct test of the effects of TBO on air traffic complexity, suggest that TBO can significantly decrease complexity perceived by air traffic controllers, albeit only when air traffic is high and when a large proportion of aircraft are flying by TBO. The lack of consistency of perceived complexity for the same air traffic controller and between different controllers in our experiments highlights the need to confirm and extend these findings in larger studies, perhaps involving optimization or recalibration of the complexity scale.

When we carefully selected 20 indicators of air traffic complexity from the literature and tested them against our simulation data, we found that they showed reasonable but not excellent fit. In fact, the goodness of fit decreased as the proportion of TBO aircraft increased. We conclude that for trajectory-based operations, researchers should apply additional, TBO-specific complexity indicators. As a first step towards identifying and validating more suitable indicators, we defined and tested 7, leading to 2 that could be combined with 4 existing factors to predict complexity much better in a TBO environment.

Future research should examine a larger number of controllers and scenarios, terminal operations, and off-nominal operations, such as inclement weather, ground problems, equipment failure, and emergency flights. Each of these experiments should aim at sample sizes comparable to this research (at least 10 controllers). This research showed that with that sample size, the effect of TBO can just barely be detected. The number of scenarios should be increased to cover all possible options as mentioned above, and that could significantly expand the scope of the research.

References

[1] *European ATM Master Plan*, 2nd ed. SESAR Joint Udertaking, Brussels, Belgium, Oct. 2012.

[2] *SESAR Milestone Deliverable D3–ATM Target Concept*, SESAR Joint Undertaking, Brussels, Belgium, 2007.

[3] *WP 4—En Route Operations*, SESAR Joint Undertaking, Brussels, Belgium, 2008.

[4] C. Meckiff, R. Chone, and J.-P. Nicolaon, "The tactical load smoother for multi-sector planning," in *Proc. 2nd FAA/EUROCONTROL ATM R&D Seminar*, Orlando, FL, USA, 1998, p. 5.

[5] B. Kirwan, R. Scaife, and R. Kennedy, "Investigating complexity factors in UK air traffic management," *Human Factors Aerosp. Safety*, vol. 1, no. 2, pp. 125–144, 2001.

[6] R. H. Mogford, J. A. Guttman, S. L. Morrow, and P. Kopardekar, "The complexity construct in air traffic control: A review and synthesis of the literature," CTA Inc., Pleasantville, NJ, USA, Tech. Rep. DOT/FAA/CT-TN95/22, 1995.

[7] S. Athènes, P. Averty, S. Puechmorel, D. Delahaye, and C. Collet, "ATC complexity and controller workload: Trying to Bridge the Gap," in *Proc. Int. Conf. Human-Comput. Interaction Aeronaut. (HCI-Aero)*, Cambridge, MA, USA, 2002.

[8] R. Christien, A. Benkouar, T. Chaboud, and P. Loubieres, "Air traffic complexity indicators & ATC sectors classification," in *Proc. 5th USA/Eur. Air Traffic Manage. Seminar*, Budapest, Hungary, Jun. 2003, pp. 1–8.

[9] A. Majumdar and W. Y. Ochieng, "The factors affecting air traffic controller workload: A multivariate analysis based upon simulation modelling of controller workload," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1788, pp. 58–69, 2002, doi: 10.3141/1788-08.

[10] P. U. Lee and T. Prevot, "Prediction of traffic complexity and controller workload in mixed equipage NextGen environments," in *Proc. Human Factors Ergonomics Soc. Annu. Meeting*, Boston, MA, USA, 2012, pp. 100–104.

[11] Wyndemere, "An evaluation of air traffic control complexity," Wyndemere Inc., Boulder, CO, USA, Tech. Rep. NAS2-14284, 1996.

[12] P. Kopardekar and S. Magyarits, "Measurement and prediction of dynamic density," in *Proc. 5th USA/Eur. Air Traffic Manage. R&D Seminar*, Budapest, Hungary, 2003, pp. 1–9.

[13] T. Radišić, D. Novak, and B. Juričić, "Development and validation of an ATC research simulator," in *Proc. Int. Conf. Air Transp. (INAIR)*, Amsterdam, The Netherlands, 2015.

[14] A. Nuić *et al.*, "Advanced aircraft performance modeling for ATM: Enhancements to the BADA model," in *Proc. 24th Digit. Avionics Syst. Conf.*, Washington, DC, USA, 2005, p. 2B4.

[15] T. Prevot and P. U. Lee, "Trajectory-based complexity (TBX): A modified aircraft count to predict sector complexity during trajectory-based operations," in *Proc. IEEE/AIAA 30th Digit. Avionics Syst. Conf. (DASC)*, Seattle, WA, USA, Oct. 2011, pp. 3A3-1–3A3-15.

[16] T. Radišić, "The effect of trajectory-based operations on air traffic complexity," Ph.D. dissertation, Faculty Transp. Traffic Sci., Univ. Zagreb, Zagreb, Croatia, 2014.

[17] B. Hilburn, "Cognitive complexity in air traffic control—A literature review," EUROCONTROL Experimental Centre, Brétigny-sur-Orge, France, EEC Note No. 04/04, Mar. 2004.

[18] P. Kopardekar, "Dynamic density—A review of proposed variables," FAA William J. Hughes Tech. Center, Adv. Concepts Branch, Internal Rep., 2000.

[19] P. Kopardekar, A. Schwartz, S. Magyarits, and J. Rhodes, "Airspace complexity measurement: An air traffic control simulation analysis," in *Proc. US/Europe 7th Air Traffic Manage. Seminar*, Barcelona, Spain, 2007.

[20] C. E. Bonferroni, *Teoria Statistica Delle Classi e Calcolo Delle Probabilita*. Florence, Italy: Libreria internazionale Seeber, 1936.

[21] O. J. Dunn, "Multiple comparisons among means," *J. Amer. Statist. Assoc.*, vol. 56, no. 293, pp. 52–64, Mar. 1961.

[22] J. W. Mauchly, "Significance test for sphericity of a normal n-variate distribution," *Ann. Math. Statist.*, vol. 11, no. 2, pp. 204–209, Jun. 1940.

[23] S. Geisser and S. W. Greenhouse, "An extension of box's results on the use of the *F* distribution in multivariate analysis," *Ann. Math. Statist.*, vol. 29, no. 3, pp. 885–891, 1958.

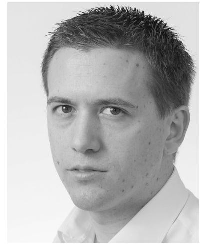
[24] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Statist. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.

[25] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, Dec. 1945.

[26] I. V. Laudeman, S. G. Shelden, R. Branstrom, and C. L. Brasil, "Dynamic density: An air traffic management metric," NASA Ames Res. Center, Moffett Field, CA, USA, Tech. Rep. NASA/TM-1998-112226, 1998.

[27] A. Masalonis, M. Callaham, and C. Wanke, "Dynamic density and complexity metrics for real-time traffic flow management," in *Proc. 5th USA/Eur. Air Traffic Manage. R&D Seminar*, Budapest, Hungary, 2003, p. 139.

[28] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 3rd ed. Hoboken, NJ, USA: Wiley, 2003.



Tomislav Radišić received the Ph.D. degree from the University of Zagreb. He is currently with the Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia, as a Research Assistant. He is also a Junior ATC Instructor with the Croatian Air Traffic Control Training Centre, where he teaches Air Navigation. His research focuses on trajectory-based operations, 4-D navigation, air traffic control simulation, air traffic complexity, and GPS signal availability.

Doris Novak received the Ph.D. degree from the University of Zagreb. He is currently an Associate Professor. He is also the Head of the Department of Aeronautics, Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia, and also the Head of the Chair of Military Aviation. His primary research interests are continuous descent approach procedures, performance based navigation, aircraft trajectory prediction, and GNSS flight procedures. He is also a Senior ATC Instructor with the Croatian Air Traffic Control Training Centre and a Senior Aviation Expert of ATM.

Biljana Juričić received the Ph.D. degree from the University of Zagreb. She played an important role in establishing and development of the Croatian Air Traffic Control (ATC) Training Centre, HUSK. The major interest of her scientific work is ATC training, airspace capacity, ATC workload, and air traffic flow management. She is currently an Assistant Professor. She is also the Head of the Chair for ATC, Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia. The main activities and responsibilities of her work are in the field of air traffic management, especially ATC.