CAPMOD Project

Final Report

Tatiana Polishchuk (LiU)
Christiane Schmidt (LiU)
Billy Josefsson (LFV)

September 2021

Trafikverket ref. 6766

CAPMOD Final Project Report

Billy Josefsson, Tatiana Polishchuk, Christiane Schmidt

Sammanfattning

Projektet Kapacitetsmodellering Stockholm RTC / Remote Tower Center) ”Capacity Modelling for Remote Controllers Workload at RTC Arlanda” (CAPMOD) syftar till kvantifiering av den operativa arbetsbelastningen för flygledaren, med fokus på fullständig och beskrivande kapacitetsmodellering, för att kunna tillgodose behoven för en effektiv framtida personalplanering och utbildning vid digital ATS (Air Traffic Service / flygtrafiktjänst) och konventionell flygplatskontrolltjänst från en fysisk tornbyggnad. Området, optimering av kompetenser med komplexa beroenden för RTC har inte beforskats tidigare och kan uppfattas som både främmande och komplext då det är nya beroenden och en annan utformning av människa–maskin interaktion en ny miljö för operativ flygtrafikledning. Spetskompetens inom logistik och modellering har i detta arbete angripit utmaningen att optimera leveransen av tjänster och därvid ta hänsyn till arbetsbelastning och påverkan från väder vilket också är ett nytt område.

Kommunikation och intresse för resultat har varit omfattande, fyra internationella publiceringar har accepterats utöver konferensbidrag. DATS konferens har under 3 år innehållit betydande rapportering från CAPMOD där Europeiska aktörer också varit aktiva vidare har arbetsmöten och presentationer hållits till NASA Ames och NTU Singapore. Viktigast har varit de nationella ”coaching sessions” som hållits 2 gånger per år för att ge verksamheten direkt tillgång till resultat och för att erhålla återkoppling. Under det treåriga projektet uppnådde vi följande:

- För det första identifierade vi en uppsättning möjliga komplexitetsfaktorer för digital ATS med hjälp av simuleringsdata från DLR (Deutsches Zentrum für Luft- und Raumfahrt): genom att analysera flygledarens subjektiva arbetsbelastning för olika situationer i en simulerad digital ATS miljö, skisserade vi händelser som på egen hand eller i samspel med en eller två andra händelser bidrar till flygtrafikledarens (ATCO) mentala arbetsbelastning och försämrade lägesmedvetenhet.
- Vi studerade flygledarens arbetsbelastning i ett konventionellt torn (Bromma) och i ett simulerat digital ATS i singel- och multipelmod (Sundsvall). Eftersom arbetsbelastningen är en ackumulerad mätning av olika stressorer (faktorer) kan en enda indikator bara delvis förklara arbetsbelastningen, medan en summa indikatorer kan ge en helhetsbild. Vi kunde identifiera faktorer som ökar arbetsbelastningsgraden, dessa faktorer är giltiga för alla ökningar av arbetsbelastningen för alla operativa flygledare både i Bromma och i Sundsvall. Regressionsanalysen för Sundsvall pekar på att summan av den genomsnittliga kommunikationen under två på varandra följande tidsperioder är en bra indikator för att bestämma arbetsbelastning, tyvärr är datamängden för liten för att dra slutliga slutsatser. Vi strävar efter att använda dessa preliminära resultat som en bas för framtida studier med större datamängder för flygplatskontrolltjänst med både konventionell och digital ATS.
- Vi analyserade ögonspårningsdata som samlats in under simuleringar av digital ATS vid Sundsvall. Vi identifierar trender i fixeringslängd och pupill-diameter som en potentiell indikator för ökad och minskad arbetsbelastning. Dessutom finner vi att en övergång från ”single” till ”multiple” mode för små flygplatser kan motverka flygledarens trötthet, monotoni och underbelastning, samtidigt som det inte ger det motsatta problemet med överbelastning. Övergången från single till multiple mode kan i sin tur bidra till ökad säkerhet (utöver effektivitet).
- Vi studerade effekterna av olika väderfenomen på flygledarens prestation. Vi identifierade uppgiftsbelastade påverkansfaktorer och motsvarande trösklar för intensiteten hos väderfenomenen på flera svenska flygplatser, vissa redan digital ATS eller är planerade för digital ATS. För att ta hänsyn till osäkerheten i väderprognosen får vi sannolikhetsväderdata från Ensemble Prediction Systems (EPS). Sedan justerade vi vår tidigare Mixed Integer Programming (MIP)-modell för RTC-personalplanering för att ta hänsyn till påverkan av väder och räkna ut en fördelning av antalet flygledare för schemaläggning av RTC -personal. Vi kvantifierade väderpåverkan genom att jämföra antalet flygledare som var nödvändiga för att producera flygplatskontrolltjänst på fem svenska flygplatser från ett digital ATS under två exempeldagar 2020, med och utan att ta hänsyn till väderhändelser, resultat visade att om inte tillräcklig hänsyn tas till väderpåverkan kan det leda till underbemanning vid RTC.

I framtida projekt strävar vi efter att fortsätta optimera produktionen i termer av kompetensgrupperingar, behörigheter och sekvensering av arbetsuppgifter genom att studera de faktorer som leder till ökad arbetsbelastning för

den operativa personalen, vilket direkt relaterar till flygsäkerhetsprestanda, uthållighet och kvalitet för den mänskliga prestationen vid flygplatskontrolltjänst för både traditionell och digitaliserad tjänst från ett RTC. Sammanfattningsvis har CAPMOD varit ett innovativt och framgångsrikt projekt som visar på logiken och argumenten för s.k. ”multiple operations” och där samhällsnyttan i form av lägre kostnader för flygtrafiktjänst blir tydlig. CAPMOD resultat ligger väl före införandeprojekten i den första vågen av digital ATS. CAPMOD visar att det finns utrymme för en mer kostnadseffektiv produktion av flygtrafiktjänst genom att gruppera kompetenser mer optimalt, inkludera arbetsbelastning och väder i bemanningen. Stöd till en sådan förändring finns bl.a. i ökad harmonisering och mer automatisering av arbetsuppgifter i kombination med ökad tillämpning av kunskaper för mänsklig prestanda, fatigue m.m.

Kunskap och modeller som utvecklats inom CAPMOD kommer att kunna tillämpas även på områdeskontrolltjänst inom flygtrafik men också inom andra transportslag. Nödvändighet att säkerhetsbevisa en förändring i det operativa baserad på CAPMOD är en bromsande aktivitet i kombination med tillgång på kompetens inom risk och säkerhetsanalys samt operativ personal för en snabb tillämpning av resultaten i CAPMOD. I ett scenario så införs delar av CAPMOD resultat på platser utanför Sverige pådrivna av kundkrav och ekonomisk reglering. Tre möjliga förslag för fortsatt arbete:

1. Robust och optimerad listning för RTC-operationer (fortsätt personaloptimering för ”multiple operations ” med hänsyn taget till arbetsbelastning/trötthet/flygsäkerhet, indikationer är att besparingar kan nå upp till 40%.
2. Väderstudier för RTC och konventionella torn (förstå hur vädret påverkar) arbetsbelastningen/arbetsbelastningen, stödverktyg kan definieras eller integreras i s.k. DITA (digital assistent).
3. En förstudie på kapacitets-/komplexitet vid en enroute central, visar att med en konservativ ansats och utan systemanpassning kan mellan 4-8% sparas för ett arbetslag per skift vid en enroute central. Därför föreslås följande områden att vidare beforskas:
   - optimering av arbetsscheman och kompetensgrupperingar
   - analys av möjlig påverkan vid ett scenario med "generisk och flexibel" flygledarkompetens, mer generiskt luftrum och ett ökat systemstöd. Vilken effekt kan förväntas? Vilka krav måste ställas på systemstöd? Vilka regulativa frågor måste besvaras?

SUMMARY (ENG)

The project Capacity Modeling for Remote Controllers Workload at RTC Arlanda (CAPMOD) targets at quantification of controller workload, focusing on complete and descriptive capacity modeling, which will serve the needs of the future personnel planning and training for delivery of air traffic services from remote controlled airports (digital Air Traffic Services) or conventional towers at airports. The area, optimization of competencies with complex dependencies for RTC has not been researched before and can be perceived as both foreign and complex as there are new dependencies and a different design of human-machine interaction a new environment for operational air traffic control. Excellence in logistics and modeling have addressed the challenges of optimizing the delivery of services and thereby take into account workload and the impact of weather, which is also a new area. Communication and interest in results have been extensive; four international publications have been accepted in addition to conference contributions. For 3 consecutive years, the DATS conference has included significant reporting from CAPMOD, where European actors have also been active, and working meetings and presentations have been held at NASA Ames and NTU Singapore. Most important have been the national "coaching sessions" which are held twice a year to give the LFV Operations direct access to results and to receive feedback on the development.

During the three-year project we achieved the following:

- First, we identified a set of possible complexity factors for multiple remote control using simulation data provided by DLR (Deutsches Zentrum für Luft- und Raumfahrt): by analyzing the ATCO subjective workload ratings of the situations in a simulated remote tower environment, we outlined events that on their own, or in co-occurrence with one or two other events drove both the air traffic controller (ATCO) mental workload and impaired situational awareness.
- We studied ATCO workload in a conventional tower (at Bromma) and in a simulated remote tower in single and multiple mode (Sundsvall). As workload is an accumulated metric of different stressors, a single indicator can only partly explain the workload, while a sum of indicators can. We were able to identify a necessary

condition for an increase in workload rating, which holds for all increases in workload rating over all ATCO ratings for both Bromma and Sundsvall. While the regression analysis for Sundsvall points to the sum of average-communication-duration weighted situations for two consecutive time periods as a good indicator for workload progression, the dataset is too small to draw final conclusions. We aim to use these preliminary results as a base for future studies with larger data sets in different tower types.

- We analysed the eye-tracking data collected during the remote tower simulation studies at Sundsvall. We identify trends in fixation duration and pupil diameter as a potential necessary indicator for workload increases and decreases. Moreover, we underline that a switch from single to multiple remote towers for small airports can counteract air traffic controller fatigue, monotony and underload, while not yielding the opposite problem of overload. The transition can in turn contribute to increased safety.
- We studied the impact of various weather phenomena on ATCO. We deduced taskload-driven impact factors and the corresponding thresholds for the intensity of the weather phenomena at several Swedish airports, which are either operated remotely or planned for remote operation. To account for the uncertainty in the weather prediction, we obtain probabilistic weather data from Ensemble Prediction Systems (EPSs). Then we adjusted our prior Mixed Integer Programming (MIP) model for RTC staff scheduling to account for uncertain impactful weather occurrences and yield a distribution for the necessary number of ATCOs for RTC staff scheduling. We quantified the impact of weather by comparing the number of controllers necessary to operate at five Swedish airports from a remote tower during two example days in 2020, with and without taking weather events into account, and showed that ignoring weather impact may lead to significant understaffing at a RTC.

In future projects we aim to continue the studies of how to group competencies in an optimal way and include factors leading to increased controllers workload but also underload which constitute a proxy to the safety and cost efficiency of remote operation. In summary, CAPMOD has been an innovative and successful project that shows the logic and arguments for the so-called "Multiple operations" and where the societal benefit in the form of lower costs for air navigation services becomes clear. CAPMOD results are well ahead of the ongoing implementation projects in the first wave of digital ATS. CAPMOD shows that there is room for a more cost-effective production of air navigation services by grouping competencies more optimally, including factors as workload and weather. Support for such a change among other things will be increased harmonization and more automation of tasks in combination with increased application of knowledge for human performance, fatigue, etc.

Knowledge and models developed within CAPMOD will also be applicable to area control services in air traffic but also in other modes of transport. Necessity to prove a change in operations based on CAPMOD is a low pace activity due to the fact that access to expertise in risk and safety analysis as well as operational personnel lacking. That is, a rapid application of the results in CAPMOD is strongly sought after but difficult to achieve. In one scenario, parts of CAPMOD's results will be introduced in locations outside Sweden driven by customer requirements and financial regulation. Three possible and plausible areas for future work:

1. Robust and optimized rostering for RTC operations (continue for staff optimization making use of multiple mode, taking into account workload/fatigue/safety performance) feasibility studies indicate benefits up to 40%
2. Weather studies for RTC and conventional towers (understanding how weather impacts taskload/workload, support tools for controllers to cope with weather)
3. Capacity / complexity studies at one enroute facility, feasibility studies show that with a conservative approach and without system adaptation, between 4-8% can be saved for one team per shift at an enroute center. Relevant further studies:
   - optimization of work schedules and competence groupings
   - analysis of possible impact in a scenario with "generic and flexible" air traffic control competence and more generic airspace with increased system support. What effect can be expected? What requirements must be placed on system support? What regulatory questions need to be answered?

A. *Roadmap*

The report is organized as follows: Section I briefly overviews the previous work related to evaluation of controller workload. Section II describes our first attempt to identify the factors influencing tower controller workload based on the simulation data provided by DLR. We present the results of the field studies at the tower Bromma and the analysis of the simulation data collected at RTC Sundsvall in single and multiple operation in Section III. In

Section IV we describe our strategy for integration of the weather impact to the staff scheduling at RTC. Section V contains the deliverables and former activities associated with the project and concludes the report.

I. State of the Art

The challenge of optimizing rosters for air traffic controllers in RTC was addressed in our previous projects KODIC I and KODIC II [5], [8], [20], [22]. In this previous work the number of Instrumental Flight Rules (IFR) flights was used as a measure of staff workload. But according to LFV Operations, IFR traffic accounts for only 40% of the workload at smaller airports, and other important aspects, which contribute to staff workload, such as ground traffic movements, bad weather conditions, seasonal variations, Visual Flight Rules (VFR), and extra traffic movements should be taken into account.

While a lot of experiments and quantitative evaluations exist for en-route traffic, this is not well researched for aerodrome control, and even less so for a remote tower control. For en-route traffic, various assessment forms of workload have been considered, see e.g. [9]. Two major approaches can be observed: subjective studies in which self-rated ATCO workload is assessed on some scale (e.g.,[26]), and objective studies that aim to find observable measures with a high correlation to an aggregation of factors that drive the complexity of an airspace (see, e.g., [37]). Pignoni and Komandur [30] recently presented a quantitative evaluation tool of cognitive workload through eye tracking. Two significant prior studies [28], [23] attempted to asses complexity in a tower environment. Please, refer to [19] for a detailed related work for the project.

II. Identification of Complexity Factors for Remote Towers

An implementation of the Remote Tower concept comes with the challenge of optimizing staff resources subject to safety requirements. To distinguish safe from unsafe assignments, the quantification of tower controller workload—which is not a new problem—needs to be reconsidered in the setting of a remote tower environment. We aim to identify the remote-operation specific complexity factors, which will be the basis of finding measures that have a high correlation to these factors that together describe the workload. We are, in particular, interested to identify complex situations that derive from the interaction of different controller tasks. Exactly these situations will be what distinguishes a workload description of a conventional tower controller from that of a remote tower controller.

A. Data

We analyze simulation data, provided by DLR (see [27], [29]). In the simulation different controllers rated the workload while monitoring multiple airports. Six teams of ATCO pairs were used for the simulation runs. All simulation scenarios had "high" traffic volume to achieve parallel movements at Erfurt and Braunschweig. The resulting dataset consists of 222 ratings for 222 situations, produced by 12 different ATCOs. On average, an ATCO rated 19 situations. Each rating consists of the following information:

- Team number
- Experimental condition, training or not
- Predefined situation number (one out of a list of nine, e.g., landing at one airport and taxiing traffic at the other)
- Subjective evaluation according to the adapted Cooper-Harper Scale
- A brief description of the problem/situation

Data preparation for the analysis consisted of a coding of the ratings based on the predefined situations and the problem description. Coding variables were created and adapted during the coding process to capture all ratings. Besides typical flight phases and connected ATCO clearances (e.g. initial call, landing, called events), conflicts, emergencies and performance problems of the ATCO (e.g. mix-up of airports) were used for coding. Finally, the coding scheme consisted of 23 variables. These variables are the initial events.

TABLE I: SUMMARY OF THE ADAPTED COOPER-HARPER SCALE BY DLR

Rating	Evaluation	Question for Evaluation
1	No problems, desirable	Is the situation solvable without major Disturbance?
2	Simple, desirable	
3	Adequate, desirable	
4	Small, but disruptive "delays"	Is the situation solvable by capacity-reducing measures?
5	Medium loss of capacity, which can be improved	
6	Very disruptive,	
	but tolerable difficulties	
7	Problems to predict development of traffic situation	Is the situation solvable if the ATCO works with a reduced situational awareness?
8	Problems in information processing	
9 9	Problems in information reception	
10	Impossible	

B. Data Collection

An adapted Cooper-Harper Scale (see the appendix of the thesis by Peters [29] for the complete scale) builds the base for the data collection—it is an adaptation of an already altered Cooper-Harper Scale that combines handling-qualities and workload to the ATC environment. The scale uses a rating from 1-10 to differentiate the impact of traffic situations on perceived handling qualities, shown in Table I. A rating of seven or higher was handled as being *critical* in terms of safety. Whilst one ATCO was controlling the traffic, the other observed the situation and was asked to assess any multiple specific situation with the adapted scale. There was a set of pre-defined situations, like two simultaneous landings. Additionally, the observer ATCO was asked to rate any situation which could only occur because of the multiple working conditions.

C. Identification of Critical Factors

In order to identify the critical complexity factors that drive the workload of a remote tower ATCO, we search for the situations at the two controlled airports in the simulation that induce a risk. We chose to analyze the data by aggregating the information w.r.t. combination of events. Combinations of events build a situation, that is, we, for example, identified all controllers that evaluated a scenario in which the two events taxi and landing appear. In the beginning, we focused on pairs of events (see Section II-D), and then considered triples of events (see Section II-E). Finally, we filtered out consequences of (simultaneous) events at two airports, and analyzed which events resulted in these problematic consequences (see Section II-F).

D. Pairs of Events

We classified the situations, defined by pairs of events by two criteria: by the average (mean) controller rating, and by the maximum controller rating. In total 65 different situations described by pairs of events where identified and compared.

1) Mean Controller Rating: The rationale behind using the mean controller rating is that a situation can be manageable or unmanageable depending on the ATCO's experience, age, and various other factors. If we aim to achieve a generic measure, we can assume an "average controller". We are particularly interested in the factors that will be problematic to this average controller. Hence, we computed the mean over all ATCO evaluations for the situation as an approximate to this average controller. The list of all event pairs with their according mean rating is shown in Figure 1. We identify all event pairs with a mean controller rating of at least 7 (shown in red color) as critical: the average controller needs to operate at least with reduced situational awareness, or even deems the situation impossible to handle. This way We identify 17 critical event pairs, where simultaneous clearance and approach, approach and conflict, clearance and go around are on top of the list.

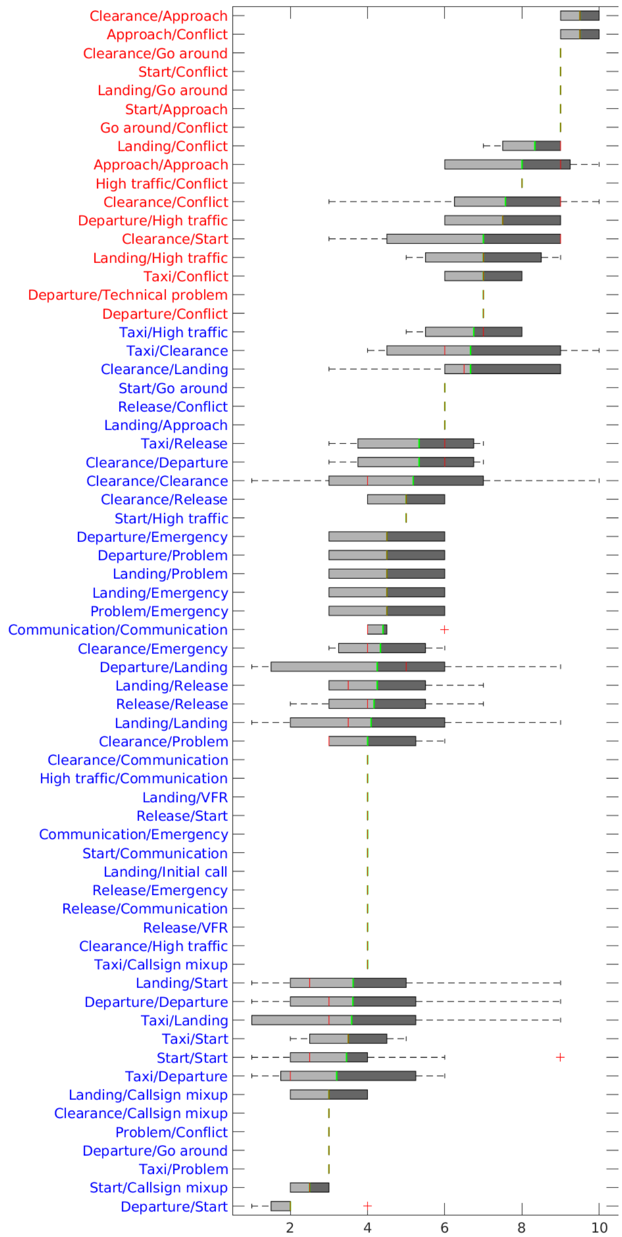


Fig. 1: Boxplot of the controller rating for the 65 event pairs ordered by **mean** controller rating, ties are broken by ordering w.r.t the maximum. The mean is shown in green, the median in red. The colors of the event pairs: red indicates a maximum controller rating of at least 7, blue otherwise.

2) Maximum Controller Rating: Using a maximum controller as the representative instead of the average controller rating could be considered as more conservative: Possibly, only a single ATCO observer experienced the criticality associated with a certain situation as very high (assigns a 7-10 to the situation), and all other ATCOs deem it solvable. However, first of all we like to identify all critical factors for the remote tower environment. Extending the list that we might have to filter out by actual correlation later on, anyhow is the approach by which failing to identify an important factor is less likely. Moreover, if we come back to our long-term application goal of such a complexity measure: if we want to ensure safe operation, we should exclude situations that will be unmanageable for any ATCO, hence, integrating all these factors is our aim in the second set of factors.

The list of all event pairs with their according maximum rating is shown in Figure 2. Obviously, the number of situations with a maximum rating of at least 7 is at least as high as the number of event pairs with a mean controller rating of at least 7. In fact, the ratio of event pairs that are deemed critical to the total number of event pairs is considerably higher (31/65) as summarized in Table II. Also the number of event pairs that obtain a maximum rating

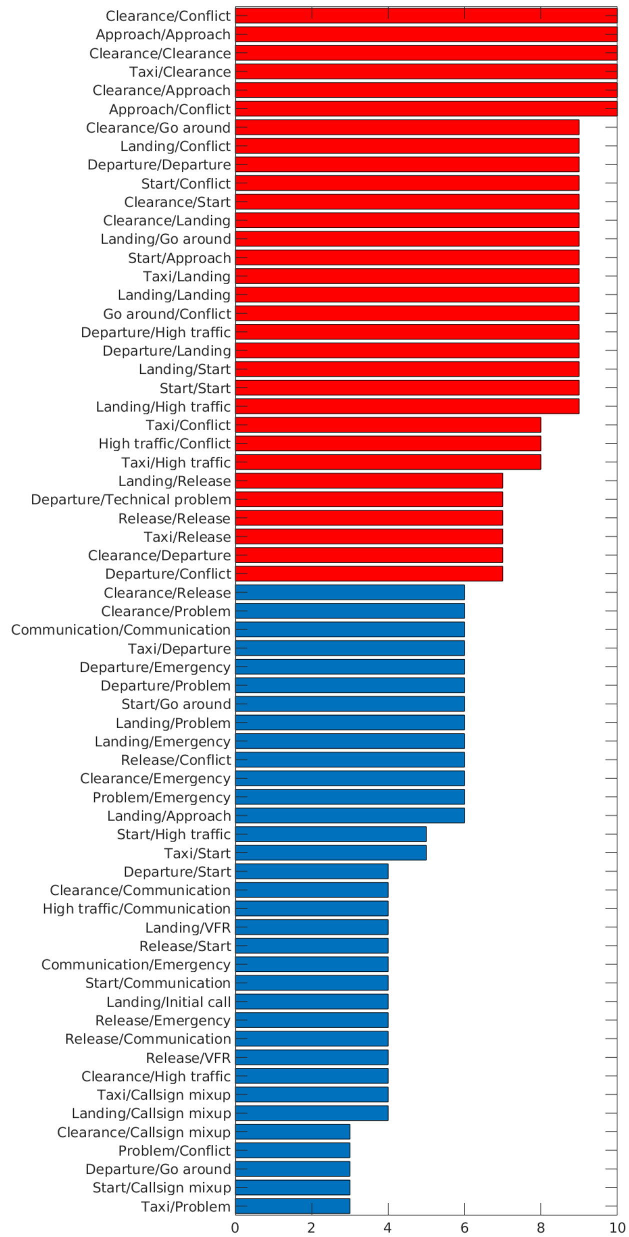


Fig. 2: Event pairs setup ordered by **max** controller rating.

TABLE II: Total number of event pairs, and share of critical event pairs

	mean controller	max controller
# identified event pairs	65	65
# event pairs with rating$\geq$ 7	17	31
share of event pairs with rating $\geq$ 7	26%	48%

of 10 (that is, are part of an impossible situation) is quite large: 5 event pairs.

E. Triples of Events

While the analysis of pairs of events gives us an idea, which factors decrease handling qualities, they often receive a higher rating when they are part of a situation with even more events. Hence, we consider the triples of events for which the rating dominates at least the rating of one of its sub-pairs. For example, for a triple of events (A,B,C) we consider the event pairs (A,B), (B,C), and (A,C), and consider it as a complicating triple if the rating of the triple (A,B,C) dominates at least one pair, e.g., (A,B)—it could dominate w.r.t. the mean or maximum rating, that is, (A,B,C) could have a mean rating of 6 and maximum rating of 9, while (A,B) has a mean rating of 5 and a maximum rating of 10, or a mean rating of 7 and a maximum rating of 8. The idea is that in this case adding an event clearly increases the complexity of the situation for the ATCO (while for a triple that does not dominate any of its sub-pairs, the intrinsic complexity seems to already stem from a combination of two factors). Of course, such dominance is in particular interesting for those triples of events that have a rating of 7 or higher w.r.t. at least one criterion, which we consider as critical triples. We consider only the UN setup.

The detailed analysis can be found in Figure 3: We only show the dominated sub-pairs and not all sub-pairs of a triple, and highlight the critical triples in orange. Most of the triples dominate at most one pair, however, there exist triples that dominate all of their sub-pairs. In the former case, only adding a third to one sub-pair increases the complexity rating, e.g., for the triple clearance/approach/conflict at a single airport, only adding an approach to the event pair clearance/conflict at a single airport will increase the complexity, while the event pairs approach/conflict and approach/clearance at a single airport already contain so much intrinsic complexity that adding the event clearance or conflict, respectively, cannot increase the controller rating. On the other hand, the triple clearance/landing/start dominates all of its sub-pairs w.r.t. the mean rating, that is, w.r.t. the average controller.

Moreover, we can observe that no critical triple contains the events emergency, call sign mix-up, communication, and problem. It is important to notice, that all critical event triples that dominate w.r.t the mean rating, dominate one sub-pair clearly, that is, here we suggest that the added event significantly increases the complexity.

F. Consequences of Events and causing factors

Finally, we filtered out consequences of (simultaneous) events at two airports, and analyzed which events contributed to these problematic consequences. The coding variables monitoring problem, small delay, mix-up of airports, switching airports, and communication problem were rated as consequences. The data set contained 96 situations with problematic consequences.

The rationale behind this analysis is that problematic consequences like a monitoring problem can be an indicator of a potentially risky, non-manageable situation, and that events that often lead to such consequences can also be considered as critical complexity factors. The results are shown in Figure 4. Some events often lead to problematic consequences, e.g., 40% of communication led to a communication problem, and—most significantly—100% of VFR traffic led to a communication problem. VFR was not one of the predefined scenario events. That is, we only know that VFR traffic was present if the ATCO mentioned it. Thus, 100% of mentions of VFR traffic coincided with a communication problem. On the other hand, go-arounds, technical problems, general problems, initial calls, outbound traffic and emergencies never were causing effect of a situation with a problematic consequence.

G. Summary

Our analysis resulted in three lists of *critical complexity factors*: first, single events that lead to critical handling qualities for at least one controller, average controller or by at least one controller—as impossible or manageable only with limited situational awareness. One main factor is the availability of relevant information. Focusing on the list of complexity factors, complexity is increased when ATCOs have to solve a traffic conflict at one airport and manage routine traffic at the second airport. The second list contains triples of events that dominate at least one of their sub-pairs and, hence, adding one of the events clearly increases the complexity of the situation for the ATCO. The third list contains factors that are likely to cause problematic consequences. Here, VFR traffic, higher traffic numbers and approaching traffic should be mentioned.

We identified a first set of possible complexity factors for multiple remote control: by analyzing the ATCO rating of situations in a simulation of an RTC environment, we identified events that on their own, or in co-occurrence with one or two other events drove both the ATCO mental workload and impaired situational awareness. We focused on factors leading to critical ratings.

Our analysis of the event pairs and triples demonstrates that there is not a single factor, but the *interplay of events* at both airports, that drives the complexity. This result pattern is known from safety research. The concept of the human performance envelope also addresses this problem [10]. Basically, not a single factor can explain performance breakdowns or critical events but the interplay of several, sometimes marginal, events. This work is a starting point for further research in the factors driving mental workload for RTC operations.

Situation	mean	min	max	Situation	mean	min	max
Clearance/Start/Callsign	3	3	3	*Taxi/Release*	5,333333333	3	7
Start/Callsign mixup	2,5	2	3	**Taxi/Landing/High traffic**	6,333333333	5	8
Taxi/Start/Start	3,5	2	5	*Taxi/Landing*	3,588235294	1	9
Start/Start	3,454545455	1	9	**Clearance/Clearance/Landing**	6,666666667	3	9
Taxi/Departure/Landing	3,5	1	6	*Clearance/Clearance*	5,181818182	1	10
Taxi/Departure	3,2	1	6	**Clearance/Landing/Landing**	6,666666667	3	9
Landing/Start/Start	3,625	1	9	*Landing/Landing*	4,090909091	1	9
Start/Start	3,454545455	1	9	**Taxi/Clearance/Clearance**	6,666666667	4	10
Taxi/Landing/Callsign	4	4	4	*Clearance/Clearance*	5,181818182	1	10
Landing/Callsign mixup	3	2	4	**Departure/Departure/Conflict**	7	7	7
Taxi/Landing	3,588235294	1	9	*Departure/Departure*	3,619047619	1	9
Start/Start/Communicati	4	4	4	**Landing/Landing/High traffic**	7	5	9
Start/Start	3,454545455	1	9	*Landing/Landing*	4,090909091	1	9
Release/Start/Start	4	4	4	**Clearance/Clearance/Start**	7	3	9
Start/Start	3,454545455	1	9	*Clearance/Clearance*	5,181818182	1	10
Landing/Release/Release	4,25	3	7	**Departure/Departure/Technical**	7	7	7
Release/Release	4,166666667	2	7	*Departure/Departure*	3,619047619	1	9
Departure/Landing/Land	4,25	1	9	**Departure/Landing/Conflict**	7	7	7
Landing/Landing	4,090909091	1	9	*Departure/Landing*	4,25	1	9
Departure/Departure/La	4,25	1	9	**Clearance/Start/Start**	7	3	9
Departure/Departure	3,619047619	1	9	*Start/Start*	3,454545455	1	9
Landing/Landing/Release	4,25	3	7	**Clearance/Departure/Conflict**	7	7	7
Landing/Landing	4,090909091	1	9	*Clearance/Departure*	5,333333333	3	7
Landing/Landing/Emerge	4,5	3	6	**Departure/Departure/High traffic**	7,5	6	9
Landing/Landing	4,090909091	1	9	*Departure/Departure*	3,619047619	1	9
Departure/Departure/Em	4,5	3	6	**Departure/Landing/High traffic**	7,5	6	9
Departure/Departure	3,619047619	1	9	*Departure/Landing*	4,25	1	9
Departure/Departure/Pro	4,5	3	6	*Landing/High traffic*	7	5	9
Departure/Departure	3,619047619	1	9	**Clearance/Clearance/Conflict**	7,571428571	3	10
Departure/Landing/Emer	4,5	3	6	*Clearance/Clearance*	5,181818182	1	10
Departure/Landing	4,25	1	9	**Taxi/High traffic/Conflict**	8	8	8
Clearance/Departure/Pro	4,5	3	6	*Taxi/High traffic*	6,75	5	8
Clearance/Problem	4	3	6	*Taxi/Conflict*	7	6	8
Landing/Landing/Proble	4,5	3	6	**Landing/Landing/Conflict**	8,333333333	7	9
Landing/Landing	4,090909091	1	9	*Landing/Landing*	4,090909091	1	9
Departure/Landing/Probl	4,5	3	6	**Clearance/Landing/Conflict**	8,333333333	7	9
Departure/Landing	4,25	1	9	*Clearance/Landing*	6,666666667	3	9
Clearance/Problem/Emer	4,5	3	6	*Clearance/Conflict*	7,571428571	3	10
Clearance/Problem	4	3	6	**Clearance/Start/Approach**	9	9	9
Clearance/Emergency	4,333333333	3	6	*Clearance/Start*	7	3	9
Clearance/Landing/Probl	4,5	3	6	**Start/Start/Approach**	9	9	9
Clearance/Problem	4	3	6	*Start/Start*	3,454545455	1	9
Clearance/Landing/Emer	4,5	3	6	**Clearance/Go around/Conflict**	9	9	9
Clearance/Emergency	4,333333333	3	6	*Clearance/Conflict*	7,571428571	3	10
Clearance/Departure/Em	4,5	3	6	**Start/Start/Conflict**	9	9	9
Clearance/Emergency	4,333333333	3	6	*Start/Start*	3,454545455	1	9
Clearance/Release/Releas	5	4	6	**Clearance/Clearance/Go around**	9	9	9
Release/Release	4,166666667	2	7	*Clearance/Clearance*	5,181818182	1	10
Start/Start/High traffic	5	5	5	**Landing/Go around/Conflict**	9	9	9
Start/Start	3,454545455	1	9	*Landing/Conflict*	8,333333333	7	9
Taxi/Start/High traffic	5	5	5	**Clearance/Start/Conflict**	9	9	9
Taxi/Start	3,5	2	5	*Clearance/Start*	7	3	9
Landing/Start/High	5	5	5	*Clearance/Conflict*	7,571428571	3	10
Landing/Start	3,625	1	9	**Clearance/Landing/Start**	9	9	9
Clearance/Departure/Lan	5,333333333	3	7	*Landing/Start*	3,625	1	9
Departure/Landing	4,25	1	9	*Clearance/Landing*	6,666666667	3	9
Taxi/Release/Release	5,333333333	3	7	*Clearance/Start*	7	3	9
Release/Release	4,166666667	2	7	**Clearance/Landing/Go around**	9	9	9
Clearance/Departure/Dep	5,333333333	3	7	*Clearance/Landing*	6,666666667	3	9
Departure/Departure	3,619047619	1	9	**Landing/Landing/Go around**	9	9	9
Clearance/Clearance/Dep	5,333333333	3	7	*Landing/Landing*	4,090909091	1	9
Clearance/Clearance	5,181818182	1	10	**Start/Approach/Approach**	9	9	9
Release/Release/Conflict	6	6	6	*Approach/Approach*	8	6	10
Release/Release	4,166666667	2	7	**Landing/Start/Conflict**	9	9	9
Landing/Landing/Approa	6	6	6	*Landing/Start*	3,625	1	9
Landing/Landing	4,090909091	1	9	*Landing/Conflict*	8,333333333	7	9
Start/Start/Go around	6	6	6	**Clearance/Approach/Approach**	9,5	9	10
Start/Start	3,454545455	1	9	*Approach/Approach*	8	6	10
Taxi/Release/Conflict	6	6	6	**Approach/Approach/Conflict**	9,5	9	10
Taxi/Release	5,333333333	3	7	*Approach/Approach*	8	6	10
Taxi/Departure/High	6	6	6	**Clearance/Clearance/Approach**	9,5	9	10
Taxi/Departure	3,2	1	6	*Clearance/Clearance*	5,181818182	1	10
Clearance/Release/Confli	6	6	6	**Clearance/Approach/Conflict**	9,5	9	10
Clearance/Release	5	4	6	*Clearance/Conflict*	7,571428571	3	10
Taxi/Clearance/Release	6	6	6				
Clearance/Release	5	4	6				

Fig. 3: Triples of events (bold) with dominated sub-pairs (italic), critical triples are highlighted in orange.

	Taxi	Clearance	Departure	Landing	Release	Start	Approach	Go around	Problem	Initial call	Technical problem	Callsign mixup	High traffic	Conflict	Commu-nication	Outbound traffic	VFR	Emergency
Monitoring problem	11.1%	0.0%	14.3%	13.6%	0.0%	0.0%	20.0%	0.0%	0.0%	0.0%	0.0%	0.0%	28.6%	0.0%	0.0%	0%	0.0%	0.0%
Small delay	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	20.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0%	0.0%	0.0%
Mix-up of airports	3.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%	0%	0.0%	0.0%
Switching airports	3.7%	0.0%	0.0%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0%	0.0%	0.0%
Communication problem	3.7%	40.9%	4.8%	6.8%	25.0%	4.5%	20.0%	0.0%	0.0%	0.0%	0.0%	10.0%	14.3%	12.5%	40.0%	0%	100.0%	0.0%

Fig. 4: Consequences and percentage of an event that lead to that consequence. The color scale indicates how many percent of an event caused a problematic consequence.

III. Validation of Controller Workload Predictors at Conventional and Remote Towers

Here we present the field studies on controller workload in a conventional and a simulated remote tower environment (in both single and multiple mode) and give a proof of concept for the validation of indicators on their workload predictability. We analyze the *number of situations* (e.g., arrivals, taxi), the *communication times* related to different situations (and use them as weights for the situations), and the *reaction times* to the Situation Present Assessment Method (SPAM) queries, as well as eye-tracking measurements (for simulated remote tower only). We show that—while the pure number of situations is not a necessary condition for an increase in workload rating—indicators that integrate the communication time related to these situations are, that is, each increase in workload rating is accompanied by an increase in these indicators.

A. Data Collection

First, we describe how the data was collected at Tower Bromma and at RTC Sundsvall, outlining the differences and similarities of the resulting datasets.

1) Bromma Tower: Bromma airport is an international airport located in Stockholm, Sweden, with ca. 4-27 movements per hours. The data collection at Bromma airport was done on March 4, 2019, using five video cameras, oriented in different directions in order to capture different views and positions. During the observation two air traffic controllers and one assistant worked in the tower. Altogether, three ATCOs were observed for four hours. The weather was snowy, and snow sweeping (with a convoy of 10-14 vehicles) appeared several times during the observation.

The workload was assessed both by observers (every five minutes) and during low traffic intensity by the two ATCOs on duty (every 15 minutes). An adapted Cooper-Harper Scale shown in Table I builds the base for the workload data collection. We used forms, which also allowed notes on the specific situation. Because the ATCOs assessed the workload only every 15 minutes, we assume that their workload was not increased due to this task. When controller shifts were exchanged data is missing for some period.

2) RTC Sundsvall: Sundsvall and Örnsköldsvik airports are regional airports located in the region of Västernorrland of Sweden. Sundsvall and Örnsköldsvik had ca. 4000 and 1700 arrivals in 2018, respectively [34]. The data collection at the Remote Tower simulator at Sundsvall was done in weeks 19/20 2019 (May 6-17). The human-in-the-loop simulations were organized by LFV researchers Lothar Meyer and Maximilian Peukert, who kindly provided the data.

For this study, we used *simulations* of both multiple operation (of Örnsköldsvik and Sundsvall airport), and single operation of Sundsvall airport at the RTM. The observation included four ATCOs, for three of which we have data for both multiple and single mode. During the simulation we had five movements in singular and six movements in multiple mode. The simulation did not include weather, and, in particular, no snow, which also means events like snow sweeping did not happen for the RT simulations.

The workload was assessed using the Instantaneous Self Assessment (ISA) scale of workload [11], [6], [18]. It uses a five-point rating scale for assessing mental workload in real time, see Table III. ATCO is prompted at regular intervals to give a value between one and five, where 1 corresponds to underutilized, and 5 to excessively busy. Exactly as the Cooper-Harper scale it can be used to compare the perceived workload. During the simulation, the regular interval was three minutes. Because two different workload scales were used for the two observations, in the last column of Table III we present an approximate way of transferring one to the other.

In addition, during the simulation trials we have collected *eye-tracking data* using Smart Eye tracking system with six infra-red cameras mounted on the work position. The system samples the head and eye with a frequency of 60Hz, which provided the following physiological and behavior-based measures of the ATCOs:

TABLE III: ISA SCALE

Rating	Workload	Spare Capacity	Description	Interpretation of CHS values
1	Underutilized	Very much	Little or nothing to do. Rather boring.	1
2	Relaxed	Ample	More time than necessary to complete the tasks. Time passes slowly.	2,3
3	Comfortable	Some	The controller has enough work to keep him/her stimulated. All tasks under control.	4,5,6
4	High	Very little	Certain nonessential tasks are postponed. Could not work at this level very long. Controller is working at the limit. Time passes quickly.	7,8,9
5	Excessive	None	Some tasks and not completed. The controller is overloaded and does not feel in control.	10

- Blink rate per minute (BR)
- Blink duration (BD)
- Pupil diameter (PD)
- Fixation duration (FD)
- Fixation frequency (FF)
- Amplitude and speed of eye closing and opening (for the closure speed, we observed a ceiling effect in the data, and we did not use it for further analysis)
- The per minute averaged eye blink duration (PERCLOS)

During all simulation runs, we had equivalent lighting conditions, hence, we do not expect an impact of lighting in our eye-tracking measurements. Consequently, we did not evaluate the fitness of our eye-tracking tool to changing light conditions, as necessary, e.g., for naval applications.

3) Situations: We defined a number of situations (a partly overlapping set of situations was first defined by Massinger and Willers [24]) that describe the current traffic situation and the resulting tasks for the ATCOs (ATs). The considered situations are:

- *Arrival:* Arriving traffic that calls the tower when turning on the ILS system or when starting a maneuver for visual approach without instrumental help.
- *Clearance:* Clearance for start, push back and landing.
- *Communication:* Communication happens during clearance, weather information, with ground traffic, and during all types of non-standard phraseology, for example, questions from the cockpit.
- *Abnormal Situation:* An abnormal situation indicates that a traffic situation becomes critical, e.g., that separation minima are undercut. This can include that arriving traffic needs to perform a go-around because of all the traffic on the runway, or birds in the area.
- *Departure:* Departing traffic that calls the tower from the gate to obtain clearance for start-up, push-back, take-off.
- *Secondary Task:* During the simulation various secondary tasks (added primary tasks, which do not affect operation, see [3]), like setting the QNH to a new value either for Örnsköldsvik or Sundsvall were requested of the ATCOs.
- *Taxi:* Aircraft that obtained clearance for taxi.

4) Communication Calls and SPAM Quiries: Apart from the number and type of situations (as defined in Subsection III-A3) and the workload rating by ATCO (and for Bromma also by an observer), we measured the length of communication calls and their purpose, and for Sundsvall we measured the reaction time of ATCOs to SPAM queries. The field study in Bromma was organized during actual operation, hence, reaction times to timed requests were not measured.

B. Tower Bromma Results

We analyze subjective workload ratings vs. the number of situations in Subsection III-B1, and the subjective workload ratings vs. various quantitative measures related to communication length in Subsection III-B3.

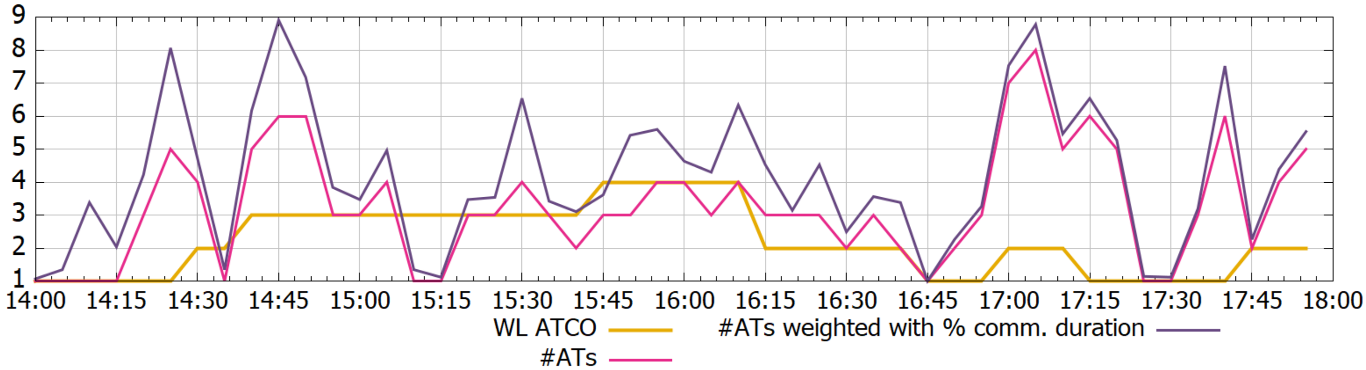


Fig. 5: Number of ATs (pink); ATs weighted with the percentage of the total communication time, see Table IV (violet); and workload assessed by ATCO (ocher) for the field study at Bromma airport.

1) Workload versus Number of ATs: The number of ATs and the workload assessed by ATCOs is shown in Figure 5 in pink and ocher, respectively. We conjecture that an increase in the workload rating is always accompanied by an increase in the number of ATs in the current or previous time period (that is, an increase in rating at 14:30 is accompanied by an increase in the number of ATs at 14:25 or 14:30). The rationale behind looking at two consecutive points in time is that more ATs in one interval may accumulate and lead to an increased workload rating at the following rating query.

The conjecture holds. However, the converse is not true: not every increase in the number of ATs leads to an increased workload rating. This suggests that an increase in the number of ATs can be a necessary, but not a sufficient indicator for increased workload.

TABLE IV: COMMUNICATION TIMES FIELD STUDY BROMMA

	Arrival	Clearance	Comm	Taxi
Average (in s)	10.04348	20.34783	11.2	10.7
Sum (in s)	231	468	448	321
Percentage	9.13%	18.49%	17.70%	12.68%
Range (in s)	6-16	6-57	4-72	5-28
	Departure	Ground	Total	
Average (in s)	11.44118	13.48	∅	
Sum (in s)	389	674	2531	
Percentage	15.37%	26.63%	100%	
Range (in s)	5-27	3-37		

2) Communication Split: Weights for Situations: Next, we analyze the time that was spent for communication in relation to different ATs. For this the duration of each radio call and its purpose was recorded. Table IV shows both the average call duration for each AT type (over all communication calls related to that AT of all ATCOs) and the sum of all radio call durations related to each AT type. Additionally, we present the latter values as percentages(see also the pie chart of Figure 6): All communication calls of all ATCOs during the observation accounted for 2531 seconds, out of which 231 seconds, or 9.13%, were related to arrivals.

When we consider the average duration of a single radio call for the different ATs, we can observe that clearances take notably more time than all other ATs. Each clearance is initiated by one party (usually it is issued by the ATCO), he/she obtains a reply by the other party, and for all airborne operations the second party then awaits a repetition of the information to confirm proper reception; this is not true for other call types, hence, the average duration of clearances is higher than that of other ATs. On the other hand, if we consider the absolute amount of time spent for radio calls related to the different ATs, clearances have an average value of 18.49%, while most time is taken up by communication to ground vehicles (26.63%), and very little time by radio calls related to arrivals (9.13%). In the total communication time to ground vehicles we clearly see the snow cleaning represented.

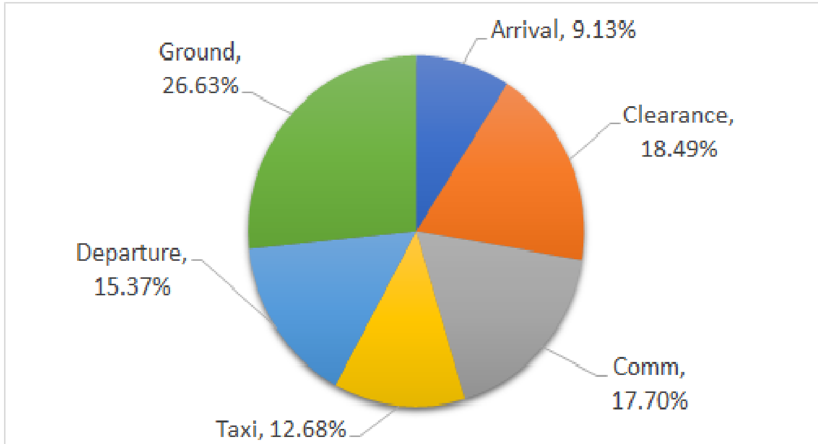


Fig. 6: Communication related to different situations in percentage of the total communication time for Bromma Tower.

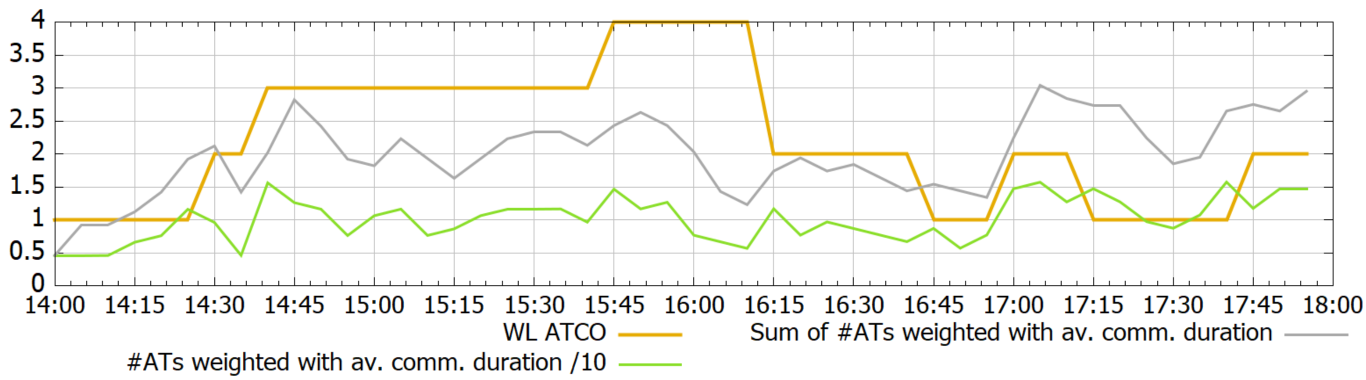


Fig. 7: ATs weighted by the average duration of radio calls related to the particular AT (green) divided by 10, the sum of this value for the current and the previous time period (gray); and the workload assessed by ATCO (ocher) for the field study at Bromma airport. See Table IV for the weights.

3) Workload versus Weighted Number of ATs: We weight the ATs with both the percentage of the *total* communication time for each AT type and the *average* communication duration for each AT type.

Percentage of the Total Communication Time. In Figure 5 we show the workload assessed by ATCOs (ocher), the number of ATs (pink) and additionally the ATs weighted with the percentage of the total communication time (violet) (as presented in Table IV). In Subsection III-B1 we conjectured that an increase in the number of ATs is a necessary condition for an increase in workload. If—instead of the pure number of ATs—we consider the ATs weighted with the percentage of the total communication time, the conjecture holds (again): We can observe that an increase in workload rating is always accompanied by an increase in the weighted ATs (with percentage of the respective AT type of the total communication time) in the current or previous time period (that is, an increase of the ATCO's workload rating at 14:40 is accompanied by an increase in the number of weighted ATs at 14:35 or 14:40).

Average Communication Duration. In addition to using the percentage of the total communication time per AT type as weights, we use the average communication duration per AT type (as shown in Table IV). Again, both the time period of workload assessment, as well as the time interval before that influence the current rating. To integrate this dependency, in Figure 7 we show—apart from the average-communication-duration weighted ATs (in green)—the sum of these values for two points in time (in gray), that is, the gray value at 15:00 equals the sum of the green values at 14:55 and 15:00.

An increase in the ATCO's workload rating is *always accompanied by an increase in at least one of*: the average-communication-duration weighted ATs (green) in the current or previous time period, and the sum of these for two time periods (gray). Hence, an increase in at least one of the average-communication-duration weighted ATs (green) in the current or previous time period and the sum of these for two time periods (gray) is a necessary condition for an increase in workload (at the later of the two time periods).

Still, an increase in *at least one of the two criteria* (the average-communication-duration weighted ATs in the current or previous time period and the sum of these for two time periods) is not a sufficient condition for an increase in workload, i.e., there exist points in time where at least one of the criteria increases, but the workload does not increase at that time period or the following time period (e.g., for the time period starting at 15:25, the average-communication-duration weighted ATs (green) increased in the previous time period, starting at 15:20, and the sum of the average-communication-duration weighted ATs for two time periods (gray) increased in the period starting at 15:25, but the workload rating did not increase in the time period starting at 15:25). Additionally, using merely the *sum of the average-communication-duration weighted ATs for two consecutive time periods* (gray) yields a necessary condition for an increase in ATCO workload rating (in the later time period).

Furthermore, we can observe that the sum of the average-communication-duration weighted ATs for *two consecutive time intervals* generally replicates the spikes and valleys in the progression of the workload rating. However, to confirm this, more observations resulting in larger data sets are needed.

Workload versus Weather. Snow sweeping with a convoy of 10-14 vehicles appeared several times during the observation. We observed 4, 5, 9 and 27 movements during the 4 hours. As a final note on the field study, we observe that the average workload rating was higher in the first three hours, during which snow sweeping occurred, than in the final hour with peak traffic (27 movements opposed to 4, 5, and 9 movements in the prior hours). More data is needed to study the influence of weather in detail. We present more results on the impact of the weather on ATCO workload in Section IV.

C. Sundsvall Results

In analogy to the Bromma study, for Sundsvall we consider the relation between subjective workload ratings and several quantitative measures, such as the number of situations and measures related to the communication length. As the simulation included remote towers *both in single and multiple mode*, we distinguish these categories. Again, we derive weights from the split of communication times over the different situations (Subsection III-C1). Additionally, we test another quantitative measure—*the reaction times*: For each SPAM query the time from the end of the query to the end of the ATCO answer was recorded, we present the results in Subsection III-C3.

1) Communication Split: Weights for ATs: We analyze the time that was spent for communication in relation to different ATs for single and multiple mode, Table V gives the average call duration for each AT type in single and multiple mode (for each ATCO and as average over all ATCOs).

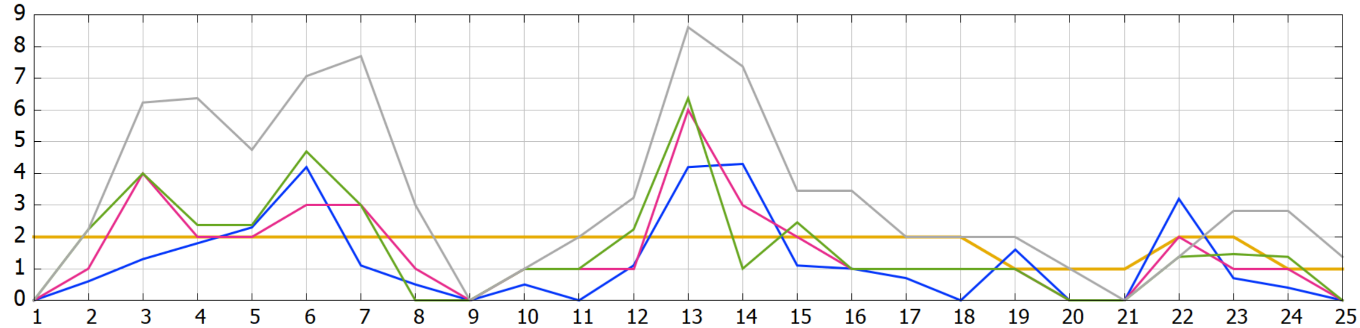
Only communication shows significantly higher values in multiple than in single mode (one-sided U-test, p-value 1.65%), the other increases are not significant. The increase in average communication times related to arrivals from multiple to single was nearly significant (one-sided U-test, p-value 7.57%). Similarly, communication for clearances shows nearly significantly higher values in multiple than in single mode (one-sided U-test, p-value 6.7%). The latter is probably caused by risk compensation behavior by the operator to avoid risk at the expense of time [25].

TABLE V: AVERAGE COMMUNICATION TIMES SIMULATION STUDIES SUNDSVALL

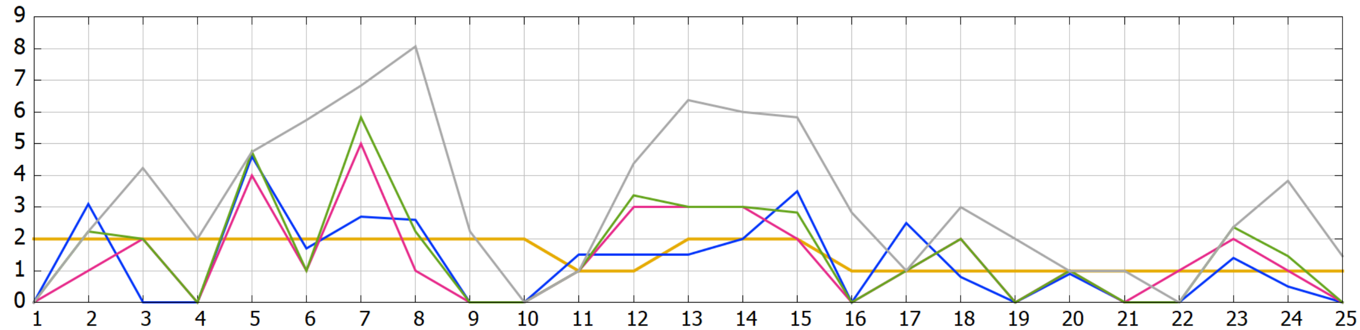
	ATCO 1 single	ATCO 1 multiple	ATCO 2 single	ATCO 2 multiple	ATCO 3 single	ATCO 3 multiple	average single	average multiple
Arrival	10.83	11.5	28.5	13.67	24	9.2	21.11	11.46
Clearance	13	22.17	13.17	13.5	12.71	25.8	12.96	20.49
Comm	8.63	13.69	10.62	11.5	9.11	12.47	9.45	12.55
Taxi	12.6	8.5	8.75	5.33	20	18.2	13.78	12.04

When using the average values (for single and multiple mode) as weights in Subsection III-C2, we first normalized the weights (that is, set the smallest value equal to 1, and then scaled the other values accordingly). For the observations in single and multiple mode, we use the average over all single and multiple average values, respectively.

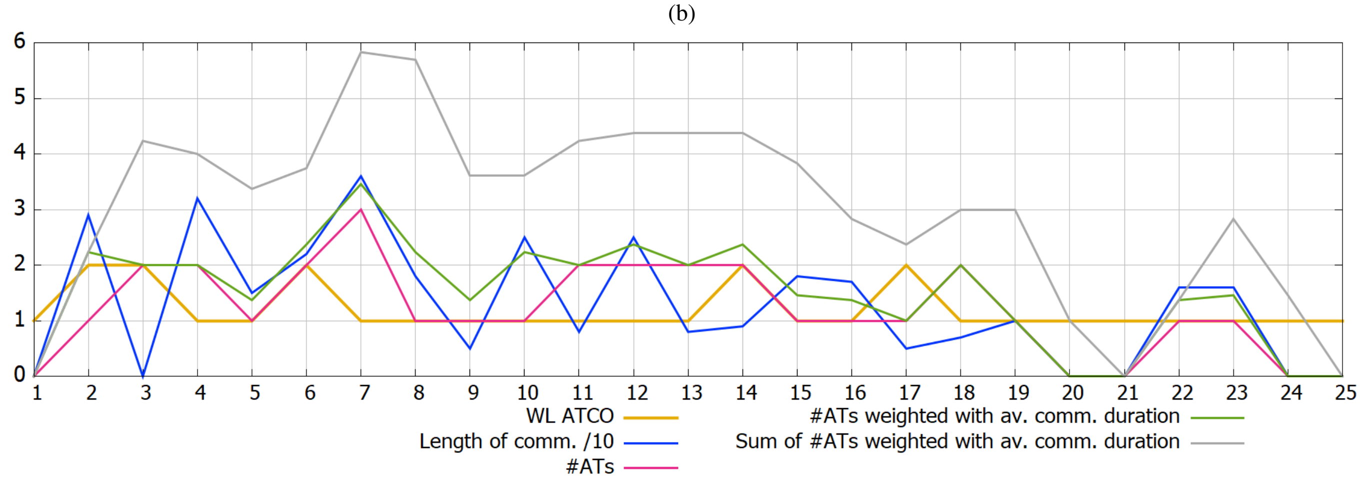
2) Workload versus Number of ATs/Weighted Number of ATs: We consider the workload versus the (weighted) number of ATs, the workload was assessed every three minutes (at 9:00, 9:03, 9:06 etc.) and the number of ATs was counted from 9:00 to 9:02:59 for time point 9:00, from 9:03 to 9:05:59 for time point 9:03 etc. Thus, we shift the workload rating numbers such that they are associated with the number of ATs up to the workload assessment.



(a)



(b)



(c)

Fig. 8: Single mode: Workload (ocher); number of ATs (pink); the length of communication at each period of observation (blue) divided by 10; average-communication-duration weighted ATs in single mode (green); and the sum of the average-communication-duration weighted ATs for two consecutive time periods (gray) in single mode for (a) ATCO 1, (b) ATCO 2, and (c) ATCO 3.

In comparison to the field study, the workload ratings for the simulation studies show smaller variation/fewer changes and long periods with the same workload assessment. This can be explained with two factors: the field study is based on the more fine-grained Cooper-Harper scale (10 values), the simulation studies are based on the ISA scale (5 values). Neither of the two studies was planned as a stress test at the boundaries of capabilities. This leaves little room for a detailed representation of the current workload using both scales, the range is, again, smaller for the ISA scale. Moreover, additional tasks like snow sweeping appeared in the field study, which led to higher variations in the taskload. The very low number of actual variations in the workload assessment hinders substantial observations.

Single Mode. We consider the workload, the number of ATs, the length of communication during each period of observation (3 mins), plus the ATs weighted by both the average communication duration in single mode and the percentage of the total communication time of specific AT types in single mode. The progression of these values for ATCO 1, 2, and 3 is shown in Figure 8(a), (b), and (c), respectively.

All ATCOs hold an endorsement for Sundsvall, i.e., they were not confronted with a new working environment in the simulation. This explains the generally low level of the workload rating (variations between 1 and 2). The rating of ATCO 3 shows larger workload variations than that of the other two ATCOs; it can be exlained by the total time of ATCO experience: 9 years for ATCO 3 (versus 20 and 41 years).

The number of ATs is not a necessary condition for an increase in workload ($\leq 43\%$ of workload rating increases were accompanied by an increase in the number of ATs). We can not observe a necessary condition for an increase in the workload rating that is valid for all ATCOs. For ATCO 1 an increase in workload rating is accompanied by an increase in all measures that take the communication time into account. For ATCO 2 each increase in the workload rating is accompanied only by an increase in the sum of the average-communication-duration weighted ATs for two consecutive time periods. For ATCO 3 an increase in workload rating is accompanied by an increase in the sum of average-communication-duration weighted ATs in all but one time period. However, if we extend the condition, and do not only include the previous, but also the following period, we obtain a necessary condition: for ATCO3 each increase in workload rating is accompanied by an increase in the average-communication-duration weighted ATs in the previous, current or following time period. The rationale behind taking the following period into account is that an ATCO anticipates later tasks, and mentally prepares for them.

Multiple Mode. We consider the workload, the number of ATs, the length of communication during each period of observation (3 mins), plus the ATs weighted by both the average communication duration in multiple mode and the percentage of the total communication time of specific AT types in multiple mode. The progression of these values for ATCO 1, 2, and 3 is depicted in Figure 9 (a), (b), and (c), respectively. In the first two time intervals ATCO 3 was stressed due to problems with the simulation equipment, hence, we start at 9:09 instead of 9:00.

ATCO 1 has the longest experience in the RTC, but an endorsement only for Sundsvall, hence, ATCO 1 was confronted with an unknown working environment, while both ATCO 2 and 3 hold endorsements for both airports. This explains the generally higher level (and higher variations) in the workload rating of ATCO 1, who—in contrast to the other ATCOs—rated some time periods with a 3, and has a "general level" at 2, while ATCO 2 has the general level at 1, and ATCO 3 fluctuates evenly between 1 and 2.

Nevertheless, we can observe a *necessary condition* for an increase in the workload rating: each increase in the workload rating (for all ATCOs) is accompanied by an increase in *at least one of the duration of communication at that time interval (blue) and the sum of average-communication-duration weighted ATs for two consecutive time periods (gray)*. This necessary condition can be compared to the necessary condition obtained for the field study. There we identified the average-communication-duration weighted ATs in the current or previous time period and the sum of these for two consecutive time points as necessary conditions. The simulation studies confirms that purely looking at the number of ATs is not enough, integrating the duration of communication yields a necessary condition. Still, in what way exactly the duration of communication constitutes a necessary condition differs slightly between the two studies, that is, either all of these need to be considered (with the formulation that an increase in at least one of these is a necessary condition for an increase in workload), or, in future research, we may aim to find a generally valid communication-length-related criterion.

For the simulation studies,—given the small data set and the human subjects—the *regression results are surprisingly good*; for ATCO 2 we yield an R^2-value of 0.33 and standard error of 0.27 for the number of ATs, an R^2-value of 0.51 and a standard error of 1.02 for the communication duration, an R^2-value of 0.39 and a standard error of 0.26 for the average-communication-duration weighted ATs, and an R^2-value of 0.53 and a standard error of 0.23 for the sum of the average-communication-duration weighted ATs for two consecutive time periods. This indicates that the sum of the average-communication-duration weighted ATs for two consecutive time periods can be a good predictor for ATCO workload. Also for ATCO 1 the R^2-value is high (0.47), but with a somewhat larger standard error of 0.42; for ATCO 3 we obtain only an R^2-value of 0.15 and a standard error of 0.47.

3) Reaction Time: Multiple vs. Singular: For various SPAM queries the reaction time of the ATCOs was measured: the time from the end of the query to the end of the ATCO answer, see [25]. The different SPAM queries were always introduced with the keyword "Question" and the categories were: SPAM clearance, SPAM wind speed, SPAM wind direction, SPAM QNH, SPAM altitude, SPAM position, and SPAM track.

Table VI and Figure 10 show the average reaction time for the three ATCOs for each SPAM query, each in single and multiple mode. For most queries we can observe that the reaction time by an ATCO in multiple mode

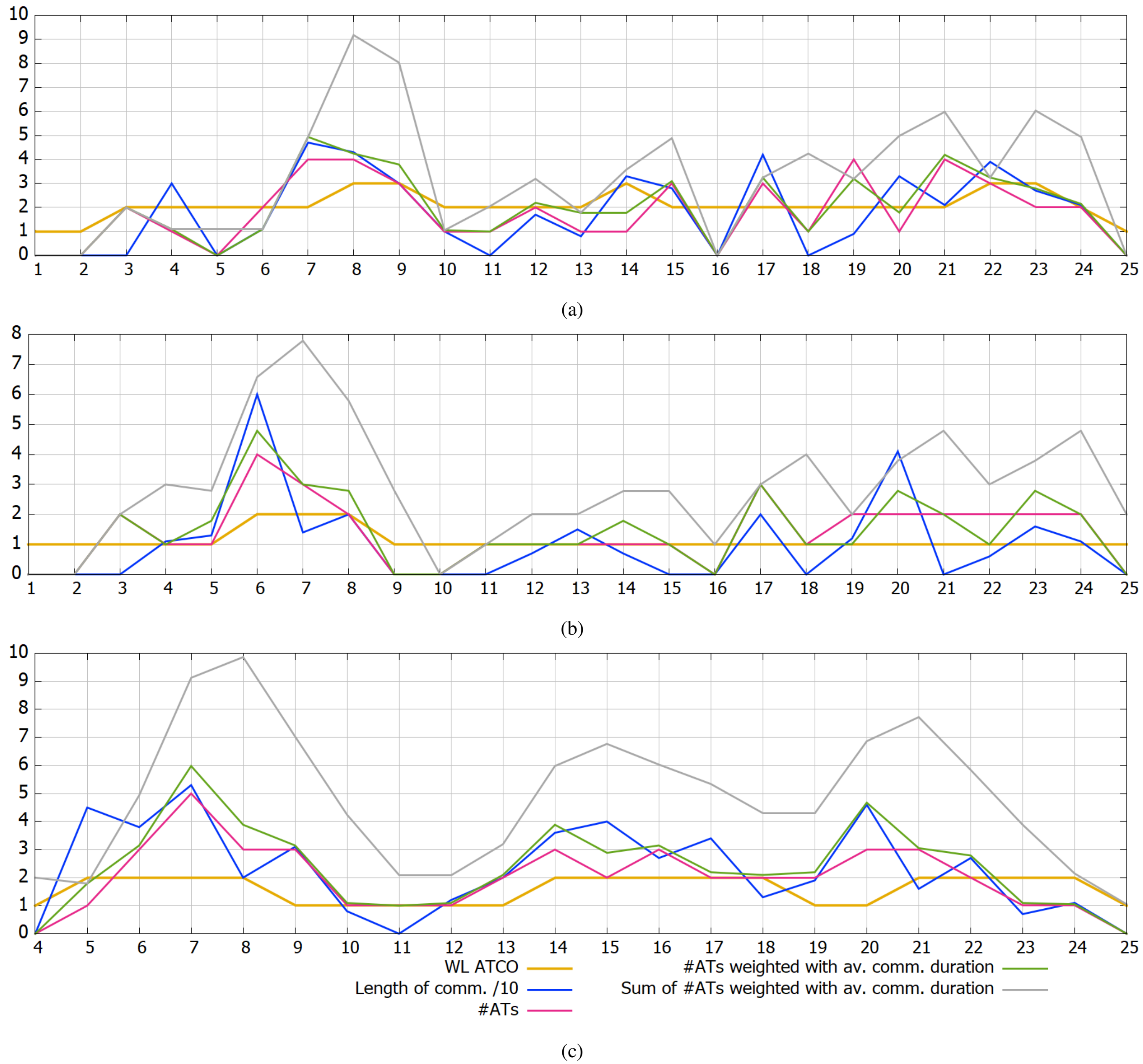


Fig. 9: Multiple mode: Workload (ocher); number of ATs (pink); the length of communication at each period of observation (blue) divided by 10; average-communication-duration weighted ATs in multiple mode (green); and the sum of the average-communication-duration weighted ATs for two consecutive time periods (gray) in multiple mode for (a) ATCO 1, (b) ATCO 2, and (c) ATCO 3.

increases in comparison with the reaction time in single mode. In multiple mode the ATCO is confronted with more tasks, hence, he/she might be less responsive—exhibit risk compensation behavior—, and this increased time can be an indicator for increased stress. On the other hand, this trend is not true for all queries and ATCOs, e.g., the reaction time of ATCO 2 for the query SPAM track reduces significantly for multiple against single mode, and reduces slightly for the queries SPAM clearance and wind speed; analogously, the reaction time for ATCO 3 for the queries SPAM position, wind speed and wind direction reduce in multiple mode. ATCO 1 had RTC experience, but holds an endorsement only for Sundsvall, hence, this ATCO was confronted with a new situation in multiple mode, while the other two ATCOs have endorsements for both controlled airports and RTC experience, which can explain the smaller increases from single to multiple mode, or even decreases in the reaction time for these (less time all over, less time is allocated for each task, while all tasks are fully under control). These observations seem

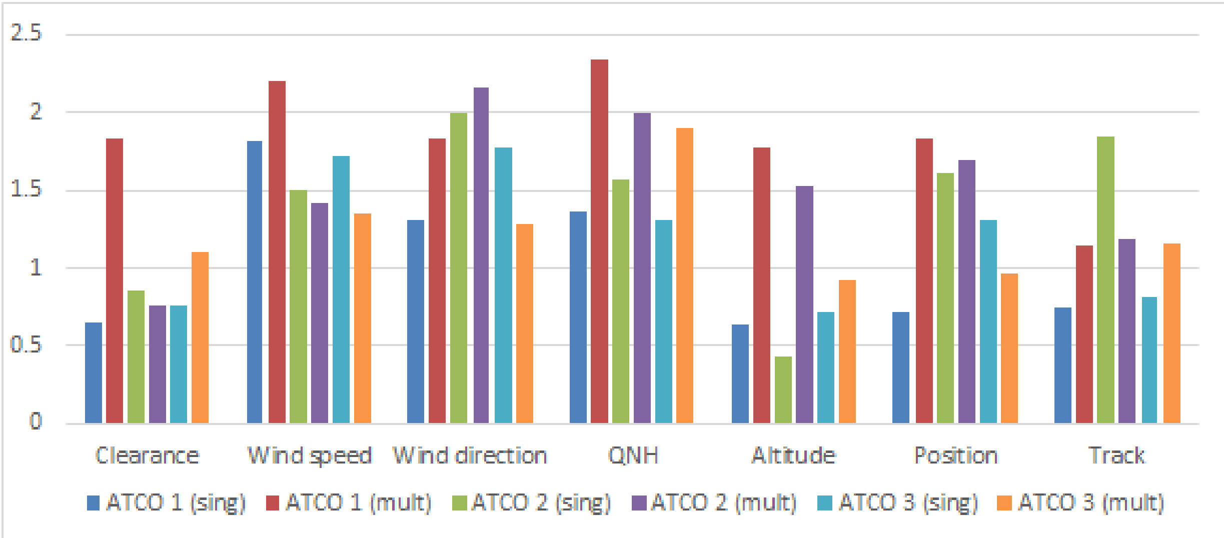


Fig. 10: Reaction times of three different ATCOs in multiple and singular mode for seven different SPAM questions (Sundsvall).

TABLE VI: REACTION TIMES SUNDSVALL

	ATCO 1 (sing)	ATCO 1 (mult)	ATCO 2 (sing)	ATCO 2 (mult)	ATCO 3 (sing)	ATCO 3 (mult)
Clearance	0.645	1.825	0.85	0.765	0.765	1.1
Wind speed	1.815	2.2	1.5	1.415	1.72	1.345
Wind direction	1.31	1.83	1.99	2.16	1.775	1.28
QNH	1.37	2.34	1.575	1.99	1.305	1.90
Altitude	0.63	1.78	0.43	1.53	0.72	0.93
Position	0.716666667	1.826666667	1.606666667	1.696666667	1.303333333	0.96
Track	0.75	1.14	1.84	1.18	0.81	1.16

to suggest that *reaction times can be a good indicator for increased stress*, which might be caused by increased workload. In the case of multiple RT the new working environment has the same effect as a stressor. Consequently, this underlines that training in a situation helps to decrease the stress of the ATCO.

4) Eye tracking Data Analysis: The eye-tracking data log was then processed using an in-house developed software that extracts the desired metrics from the log for time synchronization, data smoothing, visualization and further statistical analysis. For smoothing, we use a moving average with a time-linear sampling in 10 second steps. The moving average is defined by two parameters setting the interval range relative to the time sample.

We chose to analyze our data according to different criteria:

- **Workload:** Can we observe a connection to workload development?
- **ATCOs:** Can we observe differences between the different ATCOs?
- **Trends:** Can we observe trends in the measures over the simulation runs?

We conjecture that an increase in the workload rating is accompanied by a decrease in the *fixation duration*, we expand that by conjecturing that a decrease in workload rating is accompanied by an increase in the FD.

Workload is measured every three minutes, we have way more data points from FD even after smoothing (we smooth over the interval -180s to 0s w.r.t. the workload measurement). Hence, we consider the "trend" in the FD of a time interval: the slope of a regression line for the FD observations over the time interval. With that, we refine our conjecture:

- Each decrease in workload rating (during the period from t_i to t_{i+1}) is accompanied by a positive trend for FD in the complete three-minute interval $[t_i, t_{i+1}]$, or in the first or last 1.5 minutes of that interval.
- Each increase in workload rating (during the period from t_i to t_{i+1}) is accompanied by a negative trend for FD in $[t_i, t_{i+1}]$ or $[t_{i+1}, t_{i+2}]$ (the rationale behind considering the following period is that an ATCO anticipates later tasks and mentally prepares for them).

As an example for this hypothesis, we consider the interval between minute 27 and 30 in Figure 11: we observe a decrease in workload. According to our conjecture, this decrease should be accompanied by a positive trend for FD in the complete three-minute interval before the workload measurement, or in the first or last 1.5 minutes of that interval. The red regression line gives the FD trend for this three-minute interval. We can clearly observe the positive trend.

The hypothesis holds for the multiple mode scenarios, except for two of these simulation runs where we see one exception each—in these runs we generally observe very little workload variations. Moreover, if we have a zig-zagging pattern of the workload measurement, the succeeding three-minute period cannot be the valid criterion for workload increases (that is, workload increases must then be accompanied by a negative FD trend in the same period), to not create a contradiction with a succeeding workload decrease.

In Figure 12, we provide a full-simulation-time example of ATCO3 in multiple mode. For the workload decrease between minutes 12 and 15, we use criterion (1) with the FD trend over the first 1.5 mins of the interval only, i.e., the FD trend for $[12, 13.5]$.

We performed a *similar analysis* (necessary conditions for workload increase and decrease) using *pupil diameter*. For an examplary progression of workload and PD over a simulation run of ATCO2, see Figure 13. However, for PD, we observe three exceptions (which cannot all be assigned to single or multiple mode).

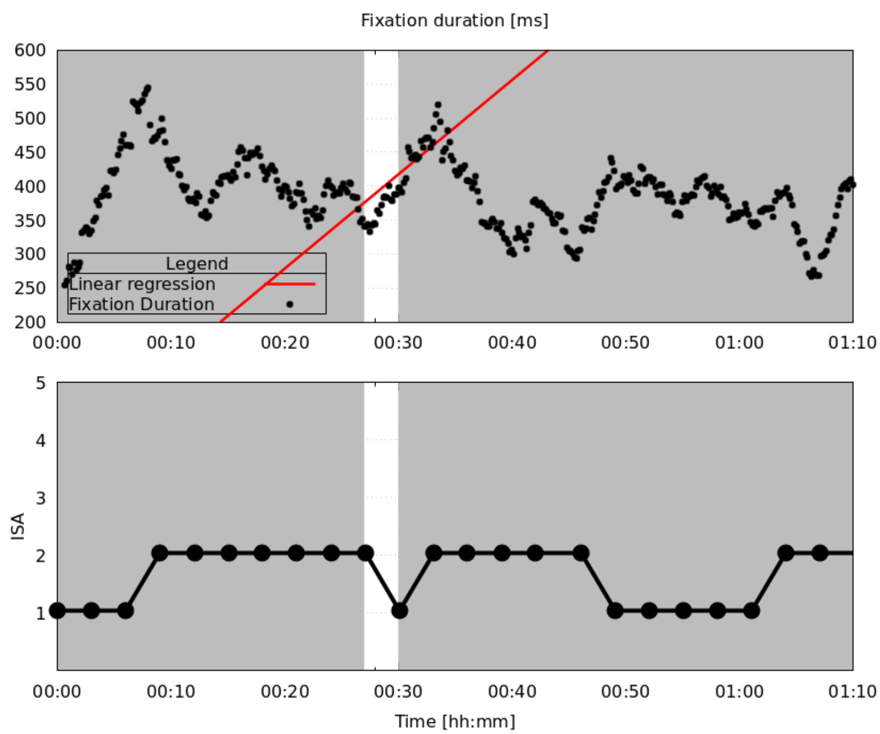


Fig. 11: Example for part (1) of our conjecture: a workload decrease between minute 27 and 30 is accompanied by a positive trend for FD in that interval. Top: FD data (to allow for visual distinction, we do not show all data points, but the smoothed data over the interval -180s to 0s w.r.t. the workload measurement) with red trend for the full interval [27:00-30:00], bottom: workload measurements according to ISA scale.

Our result suggests that *fixation duration and pupil diameter* are promising candidates for capturing the workload changes. We aim to scrutinize this in further studies with even larger data sets (in particular, for eye-tracking).

In addition, we considered a large set of behavior-based and physiological measures, but we *could not* identify any of these as promising indicators of either workload or fatigue: head yaw, head-yaw speed, PERCLOS, average eye-opening speed, amplitude of eye opening. For average eye-closing speed and amplitude of eye closing we could not draw any final conclusions as we observed a ceiling effect in our data (hence, this could still be a potentially promising measure).

For identifying a measure as promising, we used a broad set of tools: classical statistical significance tests and correlation, linear regression (e.g., for the trends of the measures over the total time of the simulation runs), detailed

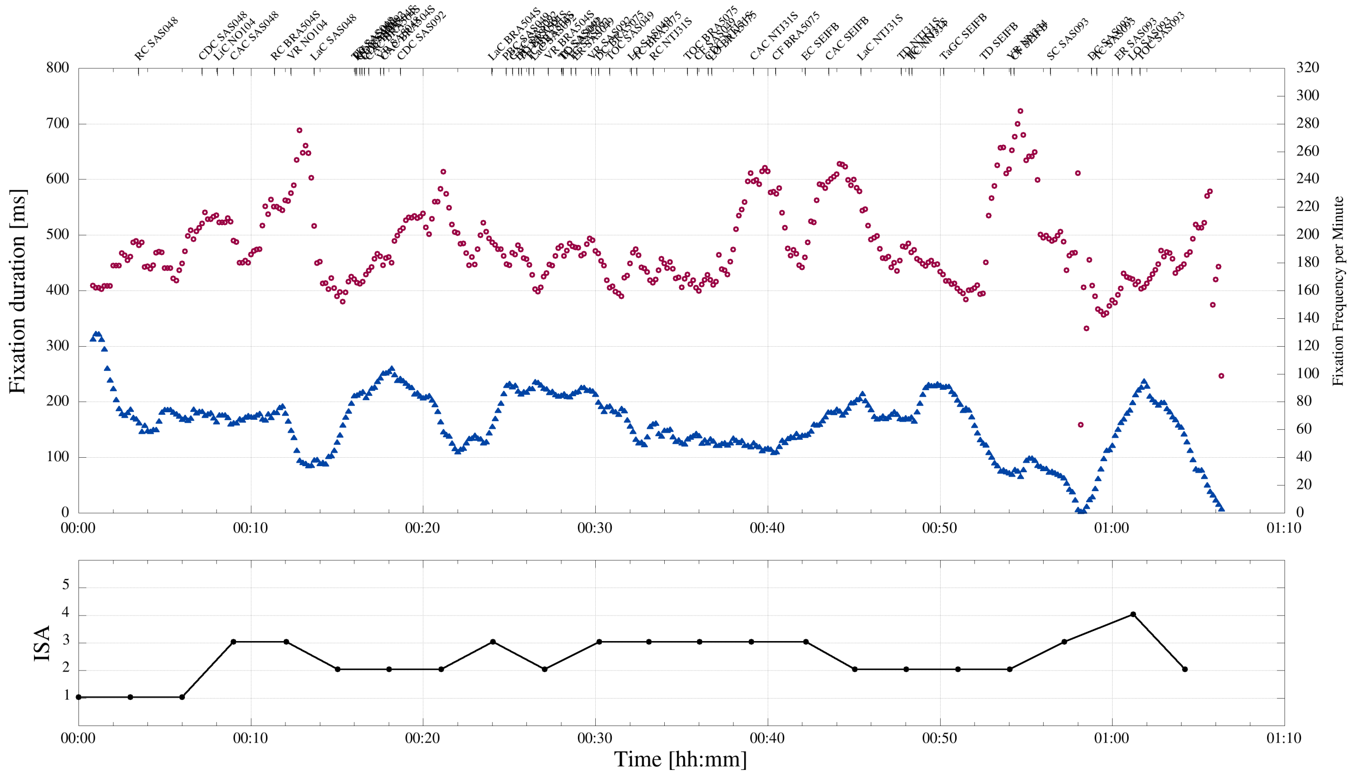


Fig. 12: Full-simulation-time example for our conjecture for ATCO3 in multiple mode. Top: FD in red circles, FF in blue triangles; bottom: workload on ISA scale.

visual analysis to conclude on potential necessary indicators—where we heavily lean on classical mathematical notation of necessary and sufficient conditions—which we then verified with our data, and comparisons of medians between modes and ATCOs. With the variety of tools we aimed to capture different types of interrelations.

D. Summary

We studied ATCO workload in a conventional tower and in a simulated remote tower in single and multiple mode. As workload is an accumulated metric of different stressors, a single indicator can only partly explain the workload, while a sum of indicators can.

We were able to identify a necessary condition for an increase in workload rating, which holds for all increases in workload rating over all ATCO ratings for both Bromma and Sundsvall: each increase in the ATCO workload rating is accompanied by an increase in at least one of

- The number of situations weighted with the percentage of the total communication time
- *The sum of average-communication-duration weighted situations*
- *The sum of average-communication-duration weighted situations for two consecutive time periods*
- *The communication duration during that time period*

In fact, just using the three criteria in italics leaves a valid statement.

All these criteria are related to the communication time in the time period of the workload assessment (and possibly the one in the previous time period). In particular, we showed that simply counting the number of situations is not a good workload indicator (it is not a necessary condition for an increase in workload rating), while taking the communication length into account leads to a necessary condition.

Of course, while identifying a necessary condition gives insights into the workload development, the identification of a sufficient criterion would be even more beneficial (that is, any increase in the criterion leads to an increase in the workload rating). For future work, we aim also for a sufficient criterion for workload rating decreases, which would altogether yield quantitative workload predictors.

While the regression analysis for Sundsvall points to the sum of average-communication-duration weighted situations for two consecutive time periods as a good indicator for workload progression, the data set is too small

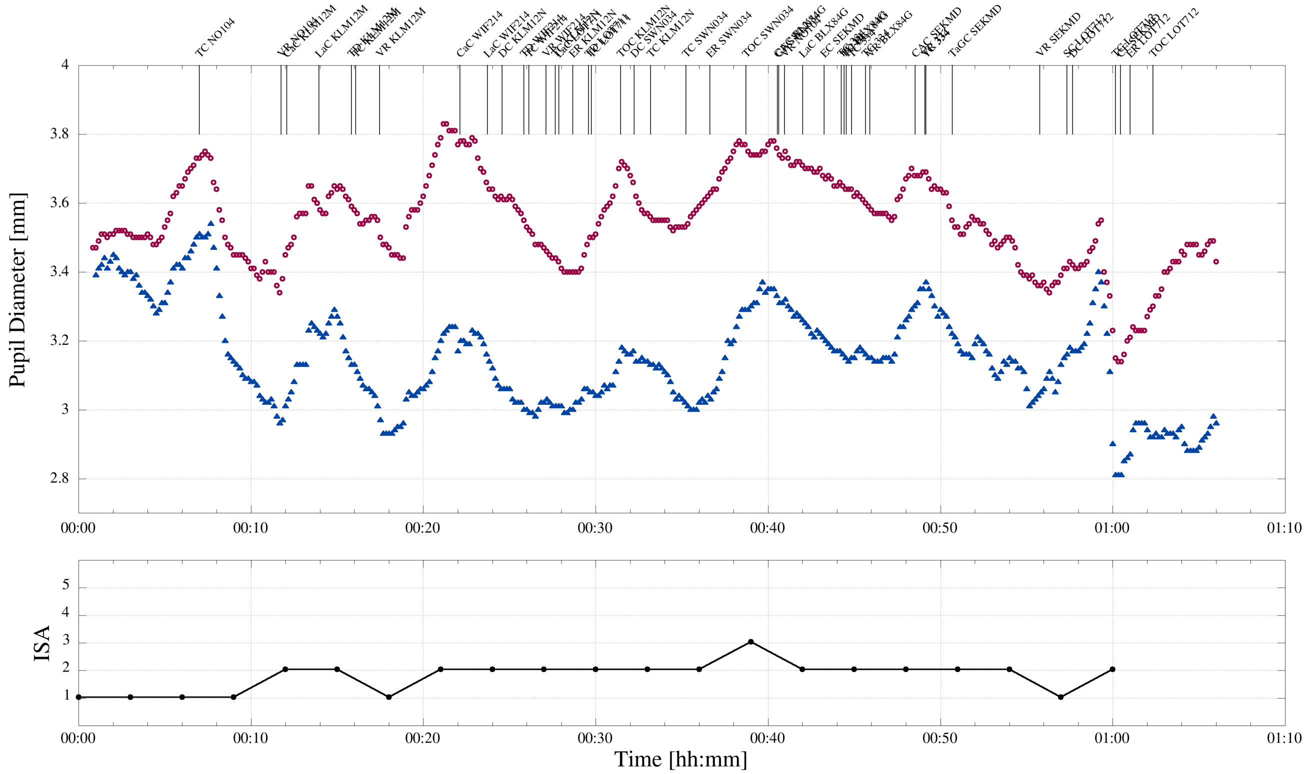


Fig. 13: Full-simulation-time example for our conjecture for ATCO2 in multiple mode. Top: PD right in red circles, PD left in blue triangles; bottom: workload on ISA scale.

to draw final conclusions. We aim to use these preliminary results as a base for future studies with larger data sets in different tower types.

IV. Integrating Weather Impact in Air Traffic Controller Shift Scheduling in Remote and Conventional Towers

Weather affects the work of air traffic controllers, however, for staff scheduling in Remote Tower Centers (RTCs) it has not been taken into account. Unions and regulatory bodies require additional safety assessment before approving multiple mode implementation. In particular, when assigning airports to ATCOs we need to ensure that no ATCO is confronted with traffic-inherent situations in the multiple airports, which would lead to an unacceptable workload for the ATCO. Such situations may stem from simultaneous movements (landings and take-offs) at two airports, or severe weather conditions in one or both airports may increase the ATCO's taskload. Weather affects the work of ATCOs through increased communication with ground traffic and pilots, through increased out-of-the-window observation, and through changes to the arrival and departure routes. For example, as described in Section III, during our field study at Bromma airport tower, we observed the influence of a severe weather was higher in the first three hours, during which regular snow sweeping with a convoy of 10-14 vehicles occurred, than in the final hour with peak traffic.

Weather disturbances have a noticeable influence on ATCOs working in a conventional tower, and they may have even higher impact on ATCOs in an RTC, in particular, when an ATCO monitors several airports simultaneously, possibly with different weather conditions. This yields a couple of research questions: How do different weather phenomena impact ATCO workload (at different airports)? How to quantify the resulting weather-induced capacity reductions? And how can we integrate this impact in RTC staff scheduling?

This work contributes to a safety assessment for multiple mode by showing that during staff scheduling we can account for weather-induced increased taskload and ensure that ATCOs do not face safety-compromising situations. Specifically, we aim to quantify the impact of weather and integrate it into optimization of ATCO work at RTCs, in particular, into an automated staff scheduling — thus revealing another important application area for weather models for aviation. We provide a proof of concept using historical weather and aircraft movement data. The goal is to use our approach as a tool to forecast ATCO staffing needs in RTCs based on weather forecasts.

A. Strategy Outline for Integrating Weather Impact in RTC Staff Scheduling

To achieve our goal of integrating weather impact in RTC staff scheduling, we implement the following steps:

(1) Identify impactful weather phenomena for each considered airport, see Section IV-E.
(2) Define threshold values for the impactful weather phenomena from (1), see Section IV-G.
(3) Obtain weather data in form of EPS, see Section IV-F.
(4) Obtain flight movements for all considered airports.
(5) Calculate a distribution of the necessary number of ATCOs for staffing based on the input from Steps (1) to (4), see Section IV-I.

B. Airports

The five Swedish airports (APs) we consider can be characterized by a couple of criteria:

- *AP1*. Small AP with low traffic, few scheduled flights per hour. Inland location north of the Arctic Circle and continental subarctic climate (Köppen climate classification Dfc, see [7]).
- *AP2*. Small regional AP with regular scheduled flights (usually open 24/7). Coastal location, Dfc, north of AP3-5.
- *AP3*. Small regional AP with regular scheduled flights. Coastal location, Dfc, north of AP5.
- *AP4*. Small regional AP with regular scheduled flights. Coastal location, Dfc, north of AP3 and AP5.
- *AP5*. Small to medium-sized AP, multiple scheduled flights per hour (usually open 24/7). Coastal location in the South of Sweden, Marine West Coast Climate.

C. ATCOs

We selected ATCOs with experience in working in a remote tower and/or with significant operational experience (as not all five airports are currently operated remotely); this way we ensured they are familiar with all weather phenomena at the selected airports.

The main goal with the interviews was to obtain the additional tasks appearing for different strengths of various weather phenomena. This mainly depends on the airport's specifics and its location and climate, but not on the subjective work assessment of specific ATCOs—that is, we target objective information. Hence, we did not aim to interview several ATCOs per airport, but to interview at least one ATCO per airport, who had long lasting experience there. One ATCO answered the interview questions for AP1-AP3; one for AP3, AP4; and one for AP5.

The three interviewed ATCOs had an average and mean experience as ATCO of 17.7 and 21 years, respectively; and an average and mean working time at the considered towers of 13 and 10 years, respectively. Two of the ATCOs have worked remotely. Since our goal is to map the additional weather-induced tasks at the five airports, experience of working remotely is not important (as several of the airports are only considered for future remote operation, but currently not operated remotely).

D. Structured Interviews

The structured interviews were performed based on a questionnaire (see [31]). Each ATCO was interviewed separately via Zoom; each interview lasted 2-3 hours (each ATCO was interviewed on one airport, and we asked them to fill in the questionnaire for other airports).

We first asked for general information on the ATCO and the ATCO career. Then, we posed questions on the traffic density and its seasonal variations at the airport: how many movements they usually observe, how many Visual-Flight-Rules (VFR) and Instrumental-Flight-Rules (IFR) movements are present on a normal day without impact by weather phenomena, and during which seasons they observe highest traffic. After this background information on ATCO and airport, we moved to weather-related questions. This included questions relating to all weather phenomena: sources for weather information, person in charge for weather updates, and frequency of weather updates, and influence of weather on staffing decisions when operating the airport from a conventional tower.

TABLE VII: Prose to Numerical Values

Prose formulation	Numerical value
no	0
rarely, not too much	0.25
sometimes, maybe, can happen, several times	0.5
often, increased, more likely, higher	0.75
yes	1
much more; yes, significantly	1.25

Thereafter, we treated different weather phenomena separately: *snow, low visibility, precipitation* (excluding snow), *wind* and *convective weather*. For each weather phenomenon we asked about its metric and usual values of this metric, and which additional ATCO or manager tasks appear in case of occurrence of that weather phenomenon.

Finally, we queried the occurrence of additional ATCO tasks in case of a light, moderate or severe occurrence of the weather phenomenon (we derived the list of possible additional tasks associated with different weather phenomena from the hazard assessment cards suggested by EUROCONTROL [12]).

Examples for the additional tasks are:

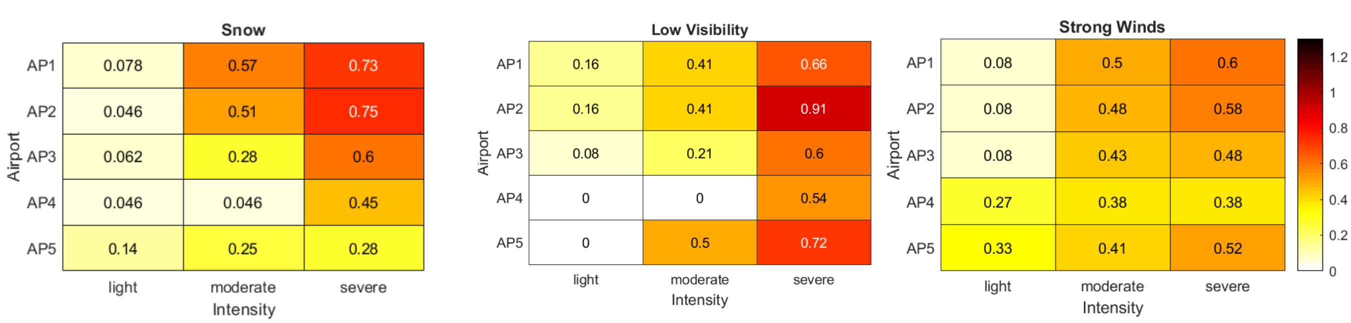
- Anticipation and condition detection
- Visual observation
- Runway closing for inspection and re-opening
- Change of departure/arrival runway
- Clearing arrivals to holding areas
- Increased coordination with the ground traffic
- Provision of information on alternate aerodromes' conditions and availability

We completed the weather-phenomenon-specific questions for each weather phenomenon with an open question on potential additional ATCO tasks the ATCO would want to add to our list.

E. Interview Results on Weather Impact

The ATCOs answered, in prose, our queries (presented in the tables in Sections 4-8 of [31]) on the occurrence of additional ATCO tasks in case of a light, moderate or severe strength of the weather phenomena. We transfer these answers to numerical values according to Table VII. Taking the average of these values for all additional ATCO tasks associated with a weather phenomenon (and for AP3 over two ATCOs' answers), we obtain *average taskload-driven impact factors* of light, moderate and severe occurrences of the weather phenomena. For easy visual differentiation of the airports, we transfer the numerical average taskload-driven impact factors to a heat value, and present the resulting impact-heat tables (see examples for snow, low visibility and strong winds in Fig. 14).

Regarding convective activity, we could not find a strong definition from the interviewed ATCOs about the difference in tasks for different intensities of convective activity (only the ATCO interviewed on AP5 was at all able to identify a difference between light, moderate, and severe thunderstorm occurrence). As a consequence, we treat thunderstorms as a binary variable, i.e, we differentiate between only two states: no convective activity and convective activity.



(a) (b) (c)

Fig. 14: Impact-heat tables for (a) snow, (b) low visibility, (c) strong winds.

From the impact-heat tables, we can clearly see that the ubiquitous weather phenomenon snow has highest impact on the northern airports AP1 and AP2, and has hardly any influence on AP5, located in the South, where it occurs rarely. Remember that we consider taskload (not workload). Hence, even though the ATCOs working at AP1 and AP2 are used to snow, its occurrence still yields additional tasks and, thus, an increased taskload. A particularly high impact of severe low visibility can be observed at AP2: a coastal airport in the North of Sweden.

These average taskload-driven impact factors allow us to differentiate the impact that different intensities of the weather phenomena have on the five airports. However, as a next step, we aim to integrate the varying impact into the staff scheduling for an RTC with AP1-AP5 in remote control. Hence, we need to decide what constitutes a *threshold* over which a weather phenomenon influences ATCO's work at an airport significantly. For that we introduce a *cutoff* value for the taskload-driven impact factor, which represents the significance of the weather event and results in the necessity to take this event into account in further analysis. We perform a sensitivity analysis on this cutoff value: we use cutoff values of 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7 and study the effect on staff scheduling for the RTC—namely, on the number of ATCOs needed to remotely control the five considered airports.

In operational staff scheduling, the *cutoff* value may be changed depending on the operator's estimate of what constitutes a *strong enough impact* to be accounted for. In Table VIII, we summarize the strength of a weather phenomenon at each airport that has an average taskload-driven impact factor of at least 0.7, 0.6, 0.5, 0.4, 0.3, and 0.2, respectively, and which, hence, is considered strong enough that it must be accounted for in staff planning.

TABLE VIII: Weather Intensity for different Average Taskload-Driven Impact Factors

		Cutoff Value for Average taskload-driven impact factor					
Airport	Intensity	≥ 0.7	≥ 0.6	≥ 0.5	≥ 0.4	≥ 0.3	≥ 0.2
AP1	Snow	severe	severe	moderate	moderate	moderate	moderate
	Low visibility	-	severe	severe	moderate	moderate	moderate
	Precipitation	severe	severe	moderate	moderate	moderate	moderate
	Strong winds	-	severe	moderate	moderate	moderate	moderate
AP2	Snow	severe	severe	moderate	moderate	moderate	moderate
	Low visibility	severe	severe	severe	moderate	moderate	moderate
	Precipitation	severe	severe	moderate	moderate	moderate	moderate
	Strong winds	-	-	severe	moderate	moderate	moderate
AP3	Snow	-	severe	severe	severe	severe	moderate
	Low visibility	-	severe	severe	severe	severe	moderate
	Precipitation	-	severe	severe	severe	severe	moderate
	Strong winds	-	-	-	moderate	moderate	moderate
AP4	Snow	-	-	-	severe	severe	severe
	Low visibility	-	-	severe	severe	severe	severe
	Precipitation	-	-	-	severe	severe	severe
	Strong winds	-	-	-	-	moderate	light
AP5	Snow	-	-	-	-	-	moderate
	Low visibility	severe	severe	moderate	moderate	moderate	moderate
	Precipitation	-	-	-	-	-	moderate
	Strong winds	-	-	severe	moderate	light	light

F. Weather Input

In this section, we describe Steps (2) and (3) from Section IV-A, related to the definition of numerical thresholds for different impactful weather phenomena and to the retrieval of probabilistic weather information from EPS.

Weather forecasts inevitably involve some level of uncertainty, which is a consequence of the chaotic nature of the atmosphere and the limited capacity to measure and model meteorological conditions. Probabilistic weather forecasts include quantitative information about this uncertainty intrinsic to meteorological predictions. One popular probabilistic weather forecasting technique is ensemble weather forecasting (EWF), which consist of generating a range of future weather possibilities. Today's trend is to use EPS, which is based on running a deterministic

Numerical Weather Prediction (NWP) model multiple times from slightly different initial conditions and with slightly perturbed weather models [36]. Typically, an EPS is a collection of 10 to 50 forecasts, referred to as members; the uncertainty information is on the spread of the members.

For the sake of illustration, the probabilistic weather information in this paper is obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis dataset [13]. The ERA5 database contains estimates of a large number of weather variables from year 1979 onwards. It covers the whole surface of the Earth, with a spacial granularity of 30 km and 137 vertical levels from the surface up to a height of 80 km. The dataset includes an uncertainty estimation for ERA5 in the form of a 10-member ensemble, which has a temporal granularity of three hours. A summary of the used ERA5 weather parameters and the correspondent variables can be found in Table IX.

TABLE IX: ERA5 parameters for impactful weather phenomena.

Weather phenomenon	ERA5 parameter	Variable
Strong winds	Instantaneous 10 metre wind gust	$i10fg$
Low visibility	Cloud base height Low cloud cover	cbh lcc
Snow	Snowfall	sf
Precipitation	Total precipitation	tp
Convective activity	Convective available potential energy Convective precipitation	$CAPE$ cp

G. Numerical Weather Thresholds

Next, we define the parameters quantifying the weather phenomena and identify the thresholds corresponding to different intensities of the weather phenomena. The selection of weather parameters and numerical weather thresholds in this work is inspired by the work of Taszarek et al. [33], where the authors use ERA5 reanalysis data to define proxies associated with hazardous weather conditions causing disruptions in European air traffic. Notice that, as mentioned in [33], the defined thresholds are only proxies of a potential impactful weather occurrence, and they cannot be considered discriminators that will perfectly distinguish between particular hazardous and non-hazardous weather events.

The chosen numerical values for the different impactful weather events of light, moderate and severe intensity for the different airports can be classified into two types: airport-specific threshold values obtained from the ATCO interviews, and general threshold values derived from literature.

a) Airport-Specific Thresholds: We have identified two weather phenomena whose threshold values were described by ATCOs in the conducted interviews, and that vary from airport to airport: strong winds and low visibility.

The weather parameter of the ERA5 reanalysis ensemble product used to identify strong winds is the instantaneous maximum wind gust at a height of ten metres above the surface of the Earth ($i10fg$). The ATCO working at airports AP1, AP2, and AP3 defined moderate wind with values between 15 knots and 25 knots for all the three airports; with this information, moderate winds in these airports are identified in the range 15 knots$\leq i10fg<$25 knots, and severe winds for $i10fg\geq$25 knots. The ATCO working at AP5 defined light wind to be more than 15 knots, moderate wind to be between 25 and 35 knots, and severe wind to be above 35 knots; the thresholds for strong winds in airports AP4 and AP5 are based on this description.

Following the work of Taszarek et al. [33], low visibility thresholds are based on the decision height defined by the ILS. Two weather parameters in the ERA5 reanalysis data, cloud base height (cbh) and low cloud cover (lcc), are used to identify this decision height. We have approximated the decision height as the cbh, with the condition that the values of lcc describe a broken (BKN) sky (over five oktas, or, equivalently, $lcc\geq$0.625) [35]. In the interviews, ATCOs at airports AP1, AP2, AP3, and AP4 identified severe low visibility at 550 m Runway Visual Range (RVR); which is the equivalent to ILS CAT I, with a decision height of 200 ft. They also described moderate low visibility as RVR values around 800 m. Taking this into account and using the decision height values for different RVRs described in [14], the cbh threshold used as a proxy for severe low visibility in airports AP1 to AP4 is $cbh\leq$200 ft (RVR$<$550 m), while moderate low visibility sits in the range 200 ft$<cbh\leq$301 ft (550 m$\leq$RVR$<$800 m). On the other hand, the ATCO in airport AP5 identified severe low visibility at 300 m (equivalent to ILS CAT II, with a decision height of 100 ft). For this last airport, we considered the described thresholds for ILS CAT I as the thresholds for light and moderate low visibility, and define a new value for severe low visibility as $cbh \leq 100$ ft (RVR<300 m).

The values of the numerical thresholds at each airport for wind and low visibility are collected in Table X.

TABLE X: Airport-dependent numerical thresholds for impactful weather phenomena

Airport	Intensity	Strong winds	Low visibility
AP1	Light	-	-
	Moderate	15 knots$\leq i10fg<$25 knots	200 ft$<cbh\leq$301 ft, $lcc\geq$0.625
	Severe	$i10fg\geq$25 knots	$cbh\leq$200 ft, $lcc\geq$0.625
AP2	Light	-	-
	Moderate	15 knots$\leq i10fg<$25 knots	200 ft$<cbh\leq$301 ft, $lcc\geq$0.625
	Severe	$i10fg\geq$25 knots	$cbh\leq$200 ft, $lcc\geq$0.625
AP3	Light	-	-
	Moderate	15 knots$\leq i10fg<$25 knots	200 ft$<cbh\leq$301 ft, $lcc\geq$0.625
	Severe	$i10fg\geq$25 knots	$cbh\leq$200 ft, $lcc\geq$0.625
AP4	Light	15 knots$\leq i10fg<$25 knots	-
	Moderate	25 knots$\leq i10fg<$35 knots	200 ft$<cbh\leq$301 ft, $lcc\geq$0.625
	Severe	$i10fg\geq$35 knots	$cbh\leq$200 ft, $lcc\geq$0.625
AP5	Light	15 knots$\leq i10fg<$25 knots	200 ft$<cbh\leq$301 ft, $lcc\geq$0.625
	Moderate	25 knots$\leq i10fg<$35 knots	100 ft$<cbh\leq$200 ft, $lcc\geq$0.625
	Severe	$i10fg\geq$35 knots	$cbh\leq$100 ft, $lcc\geq$0.625

b) General Thresholds: General thresholds are defined for snow, precipitation and convective activity events. The chosen parameter of the ERA5 ensemble product used to measure snow is the accumulated snowfall over the period of three hours, measured in meters of water equivalent; the snowfall rate (sf), measured in millimeters per hour, is obtained by evenly dividing the accumulated snowfall over these three hours. The thresholds for light, moderate and severe snowfall are obtained from the guidelines of the Society Automotive Engineers' (SAE) Ground Deicing group, also accepted by the International Civil Aviation Organization (ICAO) [17]: light snowfall falls in the range 0 to 1 mm per hour, moderate snowfall is considered to fall between 1 mm and 2.5 mm per hour, and severe snowfall is registered for values over 2.5 mm per hour.

To measure precipitation, we used the total precipitation accumulated over three hours. Similar to the case of snowfall, we divide the accumulated precipitation by the three hour time period to obtain the total precipitation rate (tp). The thresholds for light, moderate, and severe precipitation are based on the World Meteorological Organization guidelines, described by ICAO in Doc 9837 [16]: light precipitation comprises values between 0 and 2.5 mm per hour, moderate precipitation falls between 2.5 mm and 10 mm per hour, and severe precipitation exceeds 10 mm per hour.

The numerical thresholds for snowfall and precipitation are summarised in Table XI. As previously mentioned in

TABLE XI: General numerical thresholds for impactful weather phenomena

Intensity	Snow	Precipitation
Light	$0<sf\leq 1$ mmh^{-1}	$0<tp\leq 2.5$ mmh^{-1}
Moderate	1 mmh^{-1} $<sf\leq 2.5$ mmh^{-1}	2.5 mmh^{-1} $<tp\leq 10$ mmh^{-1}
Severe	$sf>2.5$ mmh^{-1}	$tp>10$ mmh^{-1}

Section IV-D, we treat convective activity as a binary variable. Following the work of Taszarek et al. in [32] and [33], thresholds for two parameters in the ERA5 reanalysis model are used as a proxy for convective initiation: convective available potential energy ($CAPE$) and convective precipitation (cp). $CAPE$ is a measurement of the potential outbreak of a thunderstorm; in particular, it is a vertical integral of the thermal buoyancy of a hypothetical air parcel that is lifted from its original vertical position [15]. For the ascending parcel considered for the calculation of $CAPE$ in the ECMWF Integrated Forecasting System (IFS), 1000 Jkg^{-1} can be considered a threshold for the potential outbreak of convective activity [13]. An additional proxy for convective activity is the occurrence of convective precipitation. Following the definition of a thunderstorm day in [32], we define a value of $cp\geq$0.075 mmh^{-1}. In summary, convective activity is identified when both conditions $CAPE\geq$1000 Jkg^{-1} and $cp\geq$0.075 mmh-1 are met.

H. Integrating Impactful Weather Periods into the MIP

Our MIP model for RTC staff scheduling is based on our prior work [21]. In this work, we add a new constraints to take care of periods during which impactful weather appears at an airport, which enforces an airport with impactful weather during an hour h to be handled in single mode during that time. We introduce a new parameter which equals 1 if airport a must be operated in single mode in period h, and the the new constraint will ensure that if an airport a must be operated in single mode in period h (because of impactful weather at a during h), an ATCO assigned to a in h may not be assigned to any other airport in h. This substitutes the old Constraint (2) in [21]:

$$\sum_{\substack{a' \in \\ A \cup A^+}} period_{i,a',h \,(\mathrm{mod}\ p)} \leq \ y_{i,h\,(\mathrm{mod}\ p)} \cdot A_{\max} - (A_{\max} - 1) s_{a,h} \cdot period_{i,a,h\,(\mathrm{mod}\ p)}; \forall i \in C,\ \forall h \in P,\ \forall a \in A$$

I. Distribution of Necessary Number of ATCOs

The probability distribution of the necessary number of ATCOs (Q) is obtained using the described MIP and EPSs as follows. For each of the M EPS ensemble members, we solve the MIP and obtain the number $Q_m, m = 1 \dots M$ of ATCOs needed. Then the probability that at most k ATCOs are needed is:

$$P(Q \leq k) = \frac{1}{M} \sum_{m=1}^{M} X_m(k), \tag{1}$$

where

$$X_m(k) = \begin{cases} 1 & \text{if} \qquad Q_m \leq k, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

J. Experimental Study: Sweden

To illustrate how the described strategy for integrating weather impact into ATCO staff scheduling can be used in practice we perform an experimental study on the example of two days of the year 2020 for AP1-AP5 in remote control. We follow Steps (1)-(5) from Section IV-A.

(1) Identify impactful weather phenomena for each considered airport: We considered AP1-5, and the impactful weather phenomena were identified as described in Subsection IV-E, Table VIII.

(2) Define threshold values for the impactful weather phenomena: We deduced threshold values for the impactful weather phenomena from Table VIII in Subsection IV-E, and the chosen values are listed in Subsection IV-G.

(3) Obtain weather data in form of EPS. We downloaded weather data from the ERA5 reanalysis database for February 2020 and July 2020 and chose two exemplary dates:

- February 16, 2020: A winter day during which four out of the five considered weather phenomena occurred: snow, low visibility, strong wind and precipitation.
- July 29, 2020: A summer day during which three out of the five considered weather phenomena occurred: low visibility, wind and precipitation.

As previously mentioned in Section IV-F, the ERA5 reanalysis ensemble presents weather data every three hours. The hourly weather variables used in this work are obtained as follows: for cumulative weather parameters (snowfall and precipitation), the accumulated quantity is divided by the length of the time period (three hours); for instantaneous weather parameters, a linear interpolation is used to obtain the intermediate hourly values.

(4) Obtain flight movement data for all airports. We obtained the number of movements per hour at each airport using FlightRadar24 historical flight data. The movement data for February 16, 2020, and July 29, 2020, is shown in Fig. 15(a) and (b), respectively. We use only hours 6-14 for February 16, 2020, and 14-22 for July 29, 2020, that is, we provide rosters for nine hours of operation for each of the days.

(5) Calculate a distribution of the necessary number of ATCOs for RTC staffing.

The following parameters are set for the experiments to reflect the safety and efficiency requirements for RTC personnel operation typical for controller shifts [21]:

Feb 16	6	7	8	9	10	11	12	13	14
AP1	0	0	1	0	1	0	1	2	1
AP2	1	1	1	1	1	2	2	2	2
AP3	0	0	0	0	0	0	0	3	0
AP4	0	0	0	0	0	0	1	1	1
AP5	1	4	0	3	3	4	2	6	4

(a)

Jul 29	14	15	16	17	18	19	20	21	22
AP1	1	1	0	1	1	0	0	0	0
AP2	1	0	3	0	2	3	2	1	2
AP3	0	1	1	0	0	0	0	0	0
AP4	0	0	0	0	0	0	0	0	0
AP5	3	4	0	0	1	4	0	0	2

(b)

Fig. 15: The number of flight movements for the five Swedish airports during nine hours of operation on (a) February 16, 2020, and (b) July 29, 2020.

(a) *Maximum number of airports assigned to one controller ($A_{\max}$)*:
The default value of the maximum number of airports assigned to a controller is set to 2.

(b) *Maximum number of movements per controller per period ($Mov_{\max}$)*:
The maximum number of movements one controller handles at the remote tower during one hour is set to 10. This conservative assumption represents a manageable workload for the ATCO.

(c) *Maximum number of controllers per airport ($C_{\max}$)*:
In this work we assume each airport is handled by one ATCO during each period of time. But in principle, for safety reasons it may be needed to assign two controllers to control one airport. Our model provides such a possibility.

(e) *Length of controller shift ($T_{\min}$ and $T_{\max}$)*: The total time a controller spends at work should be between 3 and 9 hours.

(g) *Maximum continuous time without break (z)*: Controllers should not work "in position" for longer than 4 hours without break.

(i) *Length of controller breaks ($B_{\min}$ and $B_{\max}$)*: The controller is restricted to have at lest 1 but no more than 4 hours for the breaks, which matches requirements for the lengths of controller shift and total hours "in position".

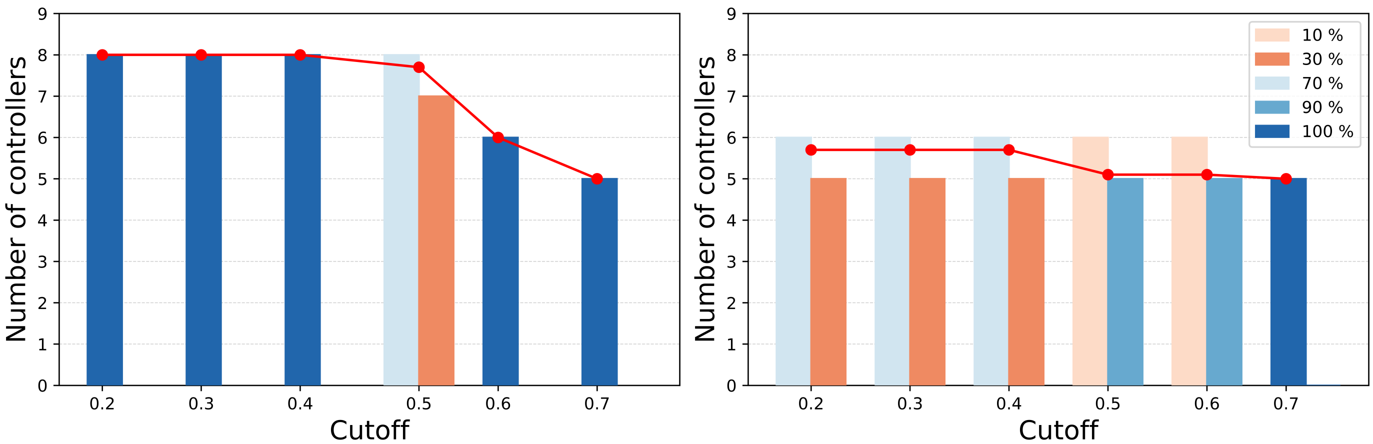
(j) *Time between the shifts ($R_{\min}$ and $R_{\max}$)*: In between two shifts an ATCO must rest for at least 2 but at most 10 hours. These parameters are based on the considered time interval of 9 hours: as we create a cyclic schedule, parameter values used for actual 24/7 operation would not create realistic solutions here. For daily or weekly schedules larger values should be used.

(k) *Period*: The length of the time period is one hour.

As described in Subsection IV-E, we use six different cutoff values (0.2, 0.3, 0.4, 0.5, 0.6, and 0.7) to perform a sensitivity analysis and study the impact on the number of ATCOs needed. We solve our MIP developed in previous work [22] for each of six cutoff values for each of ten EPS members getting as a result 60 scenarios for each day. Additionally, we—using our MIP without Constraint (IV-H)—compute the minimum number of ATCOs necessary when no weather influence is taken into account. We use Gurobi optimization software installed on a very powerful Tetralith server [4], utilizing the Intel HNS2600BPB computer nodes with 32 CPU cores, 384 GiB, provided by the Swedish National Infrastructure for Computing (SNIC). The computational time of each run of our optimization program on this powerful machine varied between 0.08 and 3.03 seconds with an average value of 0.38 seconds.

When we optimize the number of ATCOs for the 9-hour periods on both considered dates without taking weather into account, we obtain staff schedules with 5 ATCOs. Taking weather into account, we obtain different distributions of the number of necessary ATCOs dependent on the cutoff value (see also Figure 16 for bar diagrams):

- February 16, 2020
  - Cutoff 0.2: 8 ATCOs necessary with a probability of 100%
  - Cutoff 0.3: 8 ATCOs necessary with a probability of 100%
  - Cutoff 0.4: 8 ATCOs necessary with a probability of 100%
  - Cutoff 0.5: 7 and 8 ATCOs necessary with a probability of 30% and 70%, respectively
  - Cutoff 0.6: 6 ATCOs necessary with a probability of 100%
  - Cutoff 0.7: 5 ATCOs necessary with a probability of 100%



(a)

(b)

Fig. 16: Distribution of the necessary number of ATCOs with different cutoff values for (a) February 16, 2020, and (b) July 29, 2020. Red dots indicate the expected value for the necessary number of ATCOs (the red line indicates the trend of this expected value).

- July 29, 2020
  - Cutoff 0.2: 5 and 6 ATCOs necessary with a probability of 30% and 70%, respectively
  - Cutoff 0.3: 5 and 6 ATCOs necessary with a probability of 30% and 70%, respectively
  - Cutoff 0.4: 5 and 6 ATCOs necessary with a probability of 30% and 70%, respectively
  - Cutoff 0.5: 5 and 6 ATCOs necessary with a probability of 90% and 10%, respectively
  - Cutoff 0.6: 5 and 6 ATCOs necessary with a probability of 90% and 10%, respectively
  - Cutoff 0.7: 5 ATCOs necessary with a probability of 100%

K. Discussion

We can clearly observe the impact of weather and the cutoff value on the resulting distribution of the number of necessary ATCOs. If we optimize the number of ATCOs without taking weather into account, we would plan to schedule 5 ATCOs for both days. For February 16, 2020, 8 ATCOs need to be scheduled when a low taskload-driven impact factor (with cutoff value 0.2) is taken into account, while only 5 are needed for a cutoff value of 0.7. For July 29, 2020, 5 ATCOs will be sufficient only with 30% probability if we choose a low cutoff for the impact factor (0.2), but with 100% probability in case of a cutoff value of 0.7.

The choice of an appropriate cutoff value depends on the preferences of the operational manager: With a lower cutoff value a higher safety level can be achieved, as a larger set of potentially critical situations is avoided, however, this comes at the cost of using more staff members. That is, we see a clear trade-off between safety level (reflected in the chosen cutoff level) and staffing need, with the resulting higher HR cost.

We used a weather product with ten ensemble members. Consequently, our probabilities come in a step size of 10%. Thus, the lowest non-zero probability for a number of necessary ATCOs is 10%, which—depending on the desire to include even less likely scenarios—might be a reasonable value to take into account. With weather products with more ensemble members, finer granularity in the resulting distribution can be achieved. Operations might then opt to ignore a probability of, e.g., 2% (possible with at least 50 ensemble members).

While we do consider taskload-driven impact factors and while the taskload does not vary between different ATCOs, the workload, that is, the subjective stress experienced during the same tasks can vary depending on, e.g., ATCO experience, ATCO age, etc. This could be taken into account when determining the appropriate cutoff value.

L. Summary

We proposed a method to account for weather impact on ATCO work in RTC staff scheduling. We highlighted that no measures or classifications for weather impact exist, and used structured interviews with experienced ATCOs do deduce taskload-driven impact factors for five weather phenomena at five Swedish airports. We identified different sources for numerical thresholds for these impactful weather phenomena and used probabilistic weather products to obtain an ensemble of staffing solutions, from which we then derived probability distributions of the number of necessary ATCOs. To compute the ensemble of staffing solutions, we applied our prior MIP for RTC staff scheduling extended by a constraint requiring an airport with impactful weather occurrence to be operated in single mode. We presented a detailed sensitivity analysis on the cutoff value for the taskload-driven impact factor and could clearly highlight a trade-off between safety level and staffing need.

Our experiments for five Swedish airports and days with three to four weather phenomena occurring clearly show the possible impact of weather: the five ATCOs that would be scheduled taking all legal and shift-related constraints into account on both days are not always sufficient for the RTC without possibly yielding situations compromising safety due to weather. Only for a cutoff value of 0.7, the scheduling of five ATCOs will avoid what is considered as a critical situation with that value.

We highlighted the importance of developing meteorological products with longer look-ahead horizon, tailored to the needs of airports staff planning. This is particularly important for remote towers.

V. Deliverables and Activities

A. Publications:

- C. Massinger and H. Willers. En analys av den mentala arbetsbelastningen för en RATCO vid hanterbara flöden. B.S. thesis, Linköping University, 2019
- B. Josefsson, J. Jakobi, A. Papenfuss, T. Polishchuk, C.Schmidt, and L. Sedov. Identification of complexity factors for remote towers. In SESAR Innovation Days 2018.
- B. Josefsson, L. Meyer, M. Peukert, T. Polishchuk, C. Schmidt. Validation of Controller Workload Predictors at Conventional and Remote Towers. In ICRAT 2020.
- B. Josefsson, A. Lemetti, T. Polishchuk, V. Polishchuk, C. Schmidt. Integrang Weather Impact in RTC Staff Scheduling. SESAR Innovation Days (SIDs) 2020.

B. Coaching Sessions

During the project coaching session meetings took place at LFV twice a year, where we presented our results and discussed their actuality and applicability with the LFV operational experts, management and related project managers.

C. International Workshops and Research Visits

- In 2019, 2020 and 2021 we organized the international wokrshop on Digital Air Traffic Services (DATS [2]) at Linköping University, sponsored by LFV and SDATS, with the participants from many countries and organizations, e.g. EANS (Estonia), UPC Barcielona (Spain), FMI (Finland), SDATS UK, SDATS Sweden, SBworkdesign Ltd (UK), UK ATCOS, Sjofartsverket (Sweden) and Combitech (Sweden). In 2021 the workshop was organized as a virtual event and gathered attention of more than 100 participants.
- In March 2019 we visited NASA Ames research Center in San Francisco, USA, for a research meeting on Human Factors at ATM.
- We presented the summary of our results in the international live webinar "Are humans in the loop? The Future of ATCOs in Multi-Airport Digital Towers" organized by the SAAB-NTU Joint Lab in June 2021 ([1]).

References

[1] Are humans in the loop? The Future of ATCOs in Multi-Airport Digital Towers. https://www.youtube.com/watch?v=whjePq_yEJk&ab_channel=SaabNTUJointLab. Accessed on 12.08.2021.

[2] Digital Air Traffic Services Workshop (DATS), LIU-LFV, 2019-2021, Norrköping, Sweden. https://weber.itn.liu.se/~chrsc91/DATS-workshop-norrkoping/.

[3] Secondary tasks. https://tinyurl.com/y6hnd9lc. Accessed on 19/10/03.

[4] Tetralith server, NSC, Linköping University. https://www.nsc.liu.se/systems/tetralith/. Accessed on 22.09.2020.

[5] T. Andersson, P. Axelsson, J. Petersson, T. Polishchuk, V. Polishchuk, and C. Schmidt. Configuration and Planning of the Remote Tower Modules in a Remote Tower Center. In *ICRAT*, June, 2016.

[6] S. Brennen. An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. Technical Report DRA/TM/CAD5/92017, Defence Research Agency, 1992.

[7] Britannica. Köppen climate classification. https://www.britannica.com/science/Koppen-climate-classification.
[8] J. Dahlberg, T. Polishchuk, V. Polishchuk, and C. Schmidt. Stakeholder Cooperation for Improved Predictability and Lower Cost Remote Services. In *SID*, November, 2017.
[9] D. Delahaye and S. Puechmorel. Air traffic complexity: towards intrinsic metrics. In *Proceedings of the third ATM Seminar*, 2000.
[10] T. Edwards. *Human performance in air traffic control*. PhD thesis, University of Nottingham, 2013.
[11] Eurocontrol. https://ext.eurocontrol.int/ehp/?q=node/1585. accessed 2019/09/10.
[12] EUROCONTROL. Severe weather risk management survey report, 2013.
[13] European Centre for Medium-Range Weather Forecasts. ERA5: data documentation. https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+docu accessed February 2021.
[14] European Commission. Commission Regulation (EC) No 8/2008 of 11 December 2007, December 2007. http://data.europa.eu/eli/reg/2008/8/oj.
[15] P. Groenemeijer, T. Púcik, I. Tsonevsky, and P. Bechtold. An Overview of Convective Available Potential Energy and Convective Inhibition provided by NWP models for operational forecasting. (852), 11 2019.
[16] International Civil Aviation Organization. Manual on automatic meteorological observing systems at aerodromes (Doc 9873), 2011.
[17] International Civil Aviation Organization. Aerodrome meteorological observation and forecast study froup (AMOFSG). Seventh meething. Agenda Item 5: Snowfall rate thresholds for light, moderate and heavy. https://www.icao.int/safety/meteorology/amofsg/Pages/Information-Papers.aspx?meeting=AMOFSG/7, September 2008.
[18] C. Jordan. Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. Technical Report DRA/TM/CAD5/92011, Defence Research Agency, 1992.
[19] B. Josefsson, J. Jakobi, A. Papenfuss, T. Polishchuk, C.Schmidt, and L. Sedov. Identification of complexity factors for remote towers. In *SESAR Innovation Days 2018*.
[20] B. Josefsson, T. Polishchuk, V. Polishchuk, and C.Schmidt. A step towards remote tower center deployment: Optimizing staff scheduling. *Journal of Air Transportation*, 27(3), 2019.
[21] B. Josefsson, T. Polishchuk, V. Polishchuk, and C. Schmidt. Scheduling air traffic controllers at the remote tower center. In *IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, 2017.
[22] B. Josefsson, T. Polishchuk, V. Polishchuk, and C. Schmidt. Scheduling Air Traffic Controllers at the Remote Tower Center. In *DASC*, September, 2017.
[23] A. Koros, P. S. Rocco, G. Panjwani, V. Ingurgio, and J.-F. D'Arcy. Complexity in Air Traffic Control Towers: A Field Study. Part 1. Complexity Factors. Technical report, Federal Aviation Administration Technical Center, 2003.
[24] C. Massinger and H. Willers. En analys av den mentala arbetsbelastningen för en RATCO vid hanterbara flöden. B.S. thesis, Linköping University, 2019.
[25] L. Meyer, M. Peukert, B. Josefsson, and J. Lundberg. Validation of an empiric method for safety assessment of multi remote tower. In *13th USA/Europe ATM Seminar*, 2019.
[26] C. Möhlenbrink and A. Papenfuss. ATC-monitoring when one controller operates two airports research for remote tower centres. *Proc. Hum. Factors Ergon. Soc. Annu. Meet*, 55(1):76–80, 2011.
[27] C. Möhlenbrink, A. Papenfuss, and J. Jakobi. The role of workload for work organization in a remote tower control center. *Air Traffic Control Quarterly*, 20(1):5, 2012.
[28] F. Netjasov, M. Janić, and V. Tošić. Developing a generic metric of terminal airspace traffic complexity. *Transportmetrica*, 7(5):369–394, 2011.
[29] A. Papenfuss and M. Peters. HMI Laboratory Report 8: Analysis of Critical Situations at Remote Tower Operated Airports. DLR, Institut für Flugführung, Braunschweig, 2012.
[30] G. Pignoni and S. Komandur. Development of a quantitative evaluation tool of cognitive workload in field studies through eye tracking. In D. Harris, editor, *Engineering Psychology and Cognitive Ergonomics*, pages 106–122, 2019.
[31] C. Schmidt and T. Polishchuk. Questionnaire for ATCOs about the weather impact on tower operation. https://tinyurl.com/yyu26anx.
[32] M. Taszarek, J. Allen, T. Púčik, P. Groenemeijer, B. Czernecki, L. Kolendowicz, K. Lagouvardos, V. Kotroni, and W. Schulz. A climatology of thunderstorms across europe from a synthesis of multiple data sources. *Journal of Climate*, 32(6):1813 – 1837, 2019.
[33] M. Taszarek, S. Kendzierski, and N. Pilguj. Hazardous weather affecting european airports: Climatological estimates of situations with limited visibility, thunderstorm, low-level wind shear and snowfall from era5. *Weather and Climate Extremes*, 28:100243, 2020. https://www.sciencedirect.com/science/article/pii/S2212094719301100.
[34] Transportstyrelsen. Flygplatsstatistik, landningar, 2018. https://tinyurl.com/yxu9col5.
[35] U.S. Department of Transportation. Federal Aviation Administration. Advisory Circular No 00-45H. Aviation Weather Services. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_00-45H.pdf, November 2016.
[36] World Meteorological Organization. *Guidelines on Ensemble Prediction Systems and Forecasting*. WMO-No. 1091. World Meteorological Organization, 2012.
[37] E. Zohrevandi, V. Polishchuk, J. Lundberg, Å. Svensson, J. Johansson, and B. Josefsson. Modeling and analysis of controller's taskload in different predictability conditions. In *Proc. of 6th SID*, 2016.