Contrasting Safety Assessments of a Runway Incursion Scenario by Event Sequence Analysis versus Multi-Agent Dynamic Risk Modelling

Sybert H. Stroeve[1], Henk A.P. Blom[1,2], G.J. (Bert) Bakker[1]

[1]National Aerospace Laboratory NLR, Air Transport Safety Institute, Amsterdam, The Netherlands
[2]TU Delft, Faculty of Aerospace Engineering, Delft, The Netherlands
E-mail: stroeve@nlr.nl, blom@nlr.nl, bakker@nlr.nl

Abstract — **Recently we compared safety analyses for a runway incursion scenario based on an event sequence analysis, as a key exponent of a traditional risk assessment technique, versus one based on an multi-agent dynamic risk model (DRM), as an exponent of new techniques based on system complexity and variability-based accident models. We found that lower accident risk levels were assessed in the event sequence analysis and we compared various factors contributing to these differences. As the reasons of these differences were not completely understood, this paper sets forth additional analyses towards a better understanding of the relations between conflict recognition and resolution events that may occur in the runway incursion scenario and their relation to accident risk. To this end, such events were recorded in additional Monte Carlo simulations of the multi-agent DRM and a broader set of conditions was considered with agents being in or out of monitoring roles or control loops. The results of the DRM-based study uniquely make clear that the risk is not manifest from the performance of individual human operators and technical systems, nor from the sole relations between human operators and/or technical systems, but only from the totality of the performance and interactions of all human operators and technical systems in the operational context considered. In conclusion, we show that multi-agent dynamic risk modelling has considerable advantages over event sequence-based approaches.**

Keywords- Event tree; dynamic risk model; runway incursion; air traffic control; accident risk; human performance; multi-agent

I. INTRODUCTION

In complex and distributed socio-technical organizations the level of safety depends on the interactions between many entities of various types at multiple locations. The man-made disasters theory of Turner [1] gives early descriptions of how the objective of safely operating technological systems could be subverted by normal organizational processes due to unintended and complex interactions between contributory preconditions, each of which would be unlikely, singly, to defeat the established safety systems. Also Perrow [2] describes accidents as the consequence of complex interactions and tight couplings in socio-technical systems in his Normal Accident theory, stressing that given such system characteristics, multiple and unexpected interactions of failure conditions are inevitable. Building forward on the notion of normal accidents, Hollnagel [3] argues that performance in complex systems is necessarily variable due to the performance variability of its entities and the complexity of their interactions. Reasons for variability in the performance of humans include the dependency on contextual conditions, the efficiency-thoroughness trade-off in their performance and the intrinsic variability of perceptual and cognitive functions. Accidents may occur as a result of the interactions, performance variability, failures and contextual conditions of the socio-technical system.

A detailed account of complex interactions and performance variability is typically lacking in probabilistic risk assessments (PRA) of socio-technical systems by traditional event sequence-based techniques such as fault trees (FTs) and event trees (ETs). FTs represent relations between events and conditions leading to a safety-relevant situation and ETs represent relations between possible events following such a situation and the resulting consequences (e.g. accidents). They are pictorial representations of Boolean logic relations between events and they use event probabilities in PRA. The probabilities of the end events can thus be calculated straightforwardly and these end results are qualities of the same kind as the data used to obtain them: both are event probabilities. FT and ETs have been applied extensively for safety assessment in various fields, including air traffic [4][5]. An advantage of these techniques is that their structure is transparent and easy to understand. Their limitations include the difficultness to represent varieties of interdependencies between organizational entities and their dynamics, as well as the restricted evaluation of human performance by human error and conflict resolution probabilities. As such their use for risk assessment of complex socio-technical systems tends to be problematic [3][6].

In recognition of the limitations of event sequence-based techniques and in an effort to more directly address performance variability in complex socio-technical systems and the therein emergent safety risks, various methods have been developed. These developments include FRAM [3],

which pursues a qualitative analysis of safety-critical interdependencies in a functional model of an operation, STAMP [6], which uses system theoretic modelling of control loops and processes to obtain quantitative results on safety-related process variables, and TOPAZ [7], which uses multi-agent dynamic risk models (DRM) to obtain accident risk probabilities of air traffic scenarios. In multi-agent DRM accident risk is an emergent property [8][9] that is obtained by simulation of the dynamics of interacting elements in safety relevant scenarios and which uses data of these dynamics that is of a completely different nature than the accident risk. Although system complexity and performance variability-based safety assessment methods are not yet part of the standard repertoire of techniques and are being further developed, they have already been applied in several practical safety assessments, such as assessment of NASA's safety culture by STAMP [10] or risk assessment of operations of the ANSP in the Netherlands (LVNL) by TOPAZ.

To relate these two ways of thinking about the development of accidents, we performed a benchmark study for safety analyses of a particular runway incursion scenario [11][12]. In these papers we compared the results of an event sequence-based analysis with those of an assessment using a multi-agent DRM. We found that lower accident risk levels were assessed in the event sequence analysis and we compared various factors contributing to these differences. As the reasons of these differences were not completely understood, this paper goes beyond benchmarking by running additional Monte Carlo simulations in order to gain a better understanding of the relations between conflict recognition and resolution events that may occur in the runway incursion scenario and their relation to accident risk. Furthermore, this paper sets forth to contrast the probability of agents' conflict recognition and conflict resolution events with the risk effects of a broader set of hypothetical condition in the operation, where agents are in or out of monitoring roles or control loops. In this way we aim to better understand the potential of agents to restrict the risk increase in cases where the performance of other agents is affected.

This paper is organized as follows. Section II introduces the runway incursion-related safety studies. Section III describes the methods and results of the event sequence-based safety assessment. Section 0 describes the methods and results of the DRM-based safety assessment. Section V defines additional events in the MC simulations of the DRM and the results achieved. Section VI presents the risk results for conditions with agents being in or out of the monitoring and control loops. Section VII discusses the results of the event sequence and DRM approaches and their implications. Section VIII presents the conclusion of this research.

II. RUNWAY INCURSION-RELATED SAFETY STUDIES

A. Runway incursion

A runway incursion is defined by the International Civil Aviation Organization (ICAO) as "Any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and take off of aircraft" [13]. Within air traffic,

the risk of runway incursion is recognised as an important safety issue. Safety programmes such as [13][14] promote procedures and training to reduce runway incursion risk, such as following ICAO compliant procedures and naming, applying standard radiotelephony (R/T) phraseology, pilot training on aerodrome signage and markings, using standard taxi routes, etc. In addition, technology is being used and developed to reduce the likelihood and consequences of runway incursions, such as alerting systems and traffic displays. Assessment of runway incursion risk and of the potential effect of runway incursion risk reducing measures and technologies are demanding tasks, given the large number of human operators, aircraft and supporting technical systems that closely interact on the aerodrome. This complexity makes runway incursion-related safety assessments suitable candidates for comparison of the two types of accident models.

B. Safety assessments in support of taxiing operations at Amsterdam airport

The two safety assessments were done for an active runway crossing operation in good visibility conditions. The operation was proposed for crossing of runway 18C/36C at Amsterdam airport for traffic coming from and going to a new parallel runway 18R/36L. During the development of infrastructure and operational concepts for taxiing to the new runway, various risk assessment studies were done; their history is described in detail in [15]. These studies included the use of event sequences for the assessment of the risk of various safety relevant scenarios of the active runway crossing operation [16]. Having recognized the complexity of some of these scenarios, this led to the development of a multi-agent DRM for a scenario of the active runway crossing operation [17][18]. Since this DRM was developed for the same operation and considered the same set of hazards contributing to the safety relevant scenario, the models and results of these two studies provide a suitable basis for the comparison of event sequence and multi-agent DRM-based risk assessment approaches.

C. Active runway crossing operation

As the focus in this study is not on the specific results for Amsterdam airport obtained in the safety assessments, but rather on the followed lines of reasoning, in the remainder of the paper we discuss the operation and its context in generic terms. The runway considered is used for departures and has a taxiway that crosses the runway at a distance of 1000 m from the runway threshold. The visibility conditions are good.

The main human operators involved in the runway crossing operation are the pilots of the taking-off aircraft, the pilots of the taxiing aircraft, the runway controller and the ground controllers responsible for traffic on nearby taxiways. The pilots are responsible for safe conduct of the flight operations and should actively monitor for potential conflicting traffic situations. The runway controller is responsible for safe and efficient traffic handling on the runway and the runway crossings; the ground controllers are responsible for the traffic on the taxiways in the surroundings of the runway.

Aircraft may taxi across the active runway via the following procedure. First, the control over the taxiing aircraft is

transferred from a ground controller to the runway controller. The runway controller specifies a crossing clearance to the taxiing aircraft and switches off the remotely controlled stopbar. The crew of the taxiing aircraft acknowledges the clearance, initiates taxiing across the runway and reports when the taxiing aircraft has vacated the runway. After passage of the stopbar, it is automatically switched on again.

Standard communication, navigation and surveillance systems are used: communication between controllers and crews is by R/T systems, the pilots use their knowledge on the aerodrome layout and maps for taxiing, and ground radar tracking data of all aircraft and sufficiently large vehicles on the airport surface is shown on displays of the runway and ground controllers. The ATC system may generate two types of alerts to warn the runway controller: (1) a runway incursion alert for the situation that an aircraft is on the runway in front of an aircraft that has initiated to take off; (2) a stopbar violation alert for the situation that an aircraft crosses an active stopbar in the direction of the runway.

III. EVENT SEQUENCE-BASED SAFETY STUDY

In the safety assessment of [16] several safety relevant scenarios were considered for the active runway crossing operation. For the purpose of the comparison in this study, we focus on a scenario that an aircraft is taking off and a taxiing aircraft is crossing the runway while it should not; thus a runway incursion is due to the taxiing aircraft. In particular, the event sequence-based study considers that the pilot of the taxiing aircraft starts crossing without contacting the runway controller (e.g. by misunderstanding the ground controller). The ET of the runway incursion scenario, given the taxiing aircraft is crossing while it should not, considers contributions to resolution of the runway incursion conflict by the pilots of both aircraft directly or following a call by the runway controller, who may have recognized the conflict directly or via an alert. The branching points in the ET differentiate between early, medium and late recognition of the conflict by the pilots and the runway controller. This approach was chosen as a systematic means to get hold on the variety in the timing of conflict detection and resolution events by the human operators in combination with the timing of the alerts and the remaining braking distance. The outcomes of the ET specify the timing of the resolution of the conflict (early/medium/late) or the inability to timely resolve it (accident).

The parameter values of the ET are the probabilities of event occurrences. In the event sequence-based assessment, lower and upper bounds of the event probabilities were estimated by expert (controller and pilot) elicitation. Depending on the agent and the early/medium/late stage, the probabilities of the events (leading to resolution of the conflict) are in the range of 0.1 to 0.99. These probabilities must be interpreted as conditional probabilities in the ET.

The discussed ET can be condensed in a simpler aggregated ET shown in Figure 1, which neglects the resolution stage (early/medium/ late) and focuses on the contributions of no aircraft in take-off during the crossing (event Q_1^a), direct conflict recognition and resolution by the pilots (event Q_2^a),

conflict recognition by the controller independently from the alert system that leads to effective warning of the pilots and resolution of the conflict by the pilots (event Q_3^a), and conflict recognition by the controller as result of an alert that leads to effective warning of the pilots and resolution of the conflict by the pilots (event Q_4^a).
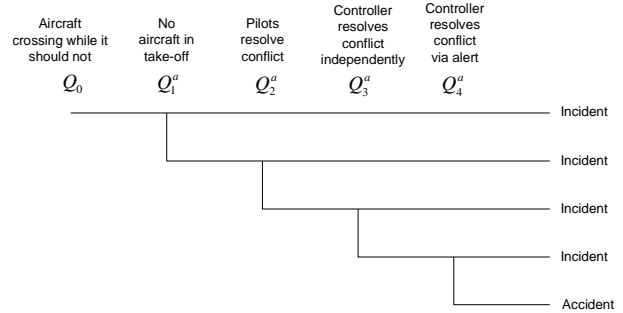

Figure 1: Aggregated ET for runway incursion scenario.

The probabilities of the events in the aggregated ET are shown in Table 1. These data include the accident risk and they reveal that in the event sequence-based safety assessment it has been assumed that the pilots have a large contribution to avoiding a collision for the runway incursion scenario (about 99.96% of the cases), the controller can only add to this independently in about half of the cases, and the controller can effectively add to the collision avoidance after an ATC alert in about 94% of the cases. An explanation of the small contribution of the controller independent from the alert system is, that as the pilot of taxiing aircraft starts crossing without contacting the runway controller, the runway controller is not very likely to timely observe the conflict by own visual monitoring. In contrast, the effectiveness of the alert system is assessed to be high as it reduces the risk by a factor 16.

TABLE 1: EVENT PROBABILITIES OF THE AGGREGATED ET.

Event		Event probability			Mean risk reduction factor
		Lower bound	Geometric mean	Upper bound	
Q_1^a	No aircraft in take-off	0.75	0.75	0.75	4.0
Q_2^a	Pilots resolve conflict	0.995	0.99961	0.99997	2600
Q_3^a	Controller resolves conflict independently	0.38	0.52	0.71	2.1
Q_4^a	Controller resolves conflict via alert	0.906	0.938	0.970	16
Accident (given aircraft crossing while it should not)		6.5E-8	2.2E-6	7.5E-5	-

IV. MULTI-AGENT DRM-BASED SAFETY STUDY

A. Multi-agent dynamic risk model

The multi-agent DRM of the runway incursion scenario is specified by a stochastic dynamic extension of the Petri net formalism [19] and is discussed in more detail in [18]. The main agents are the aircraft taking-off and taxiing, the pilots

flying of the aircraft, the runway controller and the ATC system. Key aspects of the models of these agents are highlighted next.

Taking-off Aircraft (AC-TO): The model of the taking-off aircraft represents the ground run, airborne transition and airborne climb-out phases during take-off and includes the possibility of a rejected take-off. The aircraft initiates take-off from a position near the runway threshold and it may be medium-weight or heavy-weight.

Taxiing Aircraft (AC-TX): The model of the taxiing aircraft represents aircraft movements during taxiing, including braking as a means to avoid a collision. The aircraft enters the taxiway leading to the runway crossing at a position close to the remotely controlled stopbar and its entrance time is uniformly distributed around the take-off time of AC-TO. The aircraft may be medium-weight or heavy-weight.

Surveillance (ATC subsystem): The model of the surveillance system provides position and velocity estimates for both aircraft. There is a chance that the surveillance system is not available, resulting in track loss. Surveillance data is used by the ATC alert system.

Alerts (ATC subsystem): A stopbar violation alert (SVA) becomes active if the surveillance data indicate that AC-TX has passed an active stopbar. A runway incursion alert (RIA) becomes active if the surveillance data indicate that AC-TX is within a critical distance of the runway centre-line and AC-TO has exceeded a velocity threshold in front of the runway crossing. There is a chance that the alerts are not well functioning.

R/T (ATC subsystem): The model for the R/T system between the runway controller and the aircraft crews accounts for the communication system of the aircraft, the communication system of the controller, the tower communication system and the frequency selection of the aircraft communication system. The nominal status of these communication systems accounts for direct non-delaying communication. The model accounts for the chance of delay or failure of the communication systems.

Pilot flying of the Taking-off Aircraft (PF-TO): The model for the performance of PF-TO accounts for performance of tasks such as auditory monitoring, visual monitoring, crew coordination, aircraft control, and conflict detection and reaction. The model includes dynamic representations of situation awareness about AC-TO, AC-TX and controller calls, a cognitive control mode of the pilot and task scheduling by the pilot. Initially, PF-TO is aware that take-off is allowed and initiates a take-off. During the take-off, PF-TO visually monitors the traffic situation on the runway at stochastically distributed times. PF-TO may detect a conflict if AC-TX is observed to be within a critical distance of the runway or due to an R/T call by the runway controller (ATCo-R). Following conflict detection, PF-TO starts a collision avoiding braking action if it is expected that braking will stop AC-TO in front of AC-TX; otherwise it continues and may fly over AC-TX.

Pilot Flying of Taxiing Aircraft (PF-TX): The model structure of PF-TX is similar to that of PF-TO. In the conflict scenario considered, PF-TX intends to continue taxiing on a regular taxiway (whereas actually the aircraft is on the runway crossing). During taxiing PF-TX visually monitors the traffic situation at stochastically distributed times. PF-TX may detect a conflict if AC-TX is within a critical distance of the runway, AC-TO approaches towards AC-TX and the speed of AC-TO exceeds a threshold value, or due to an R/T call of ATCo-R. Following conflict detection, PF-TX starts a collision avoiding braking action unless AC-TX already is within a critical distance of the runway centre-line; otherwise it continues and may pass the runway in front of AC-TO.

Runway Controller (ATCo-R): The model for the performance of ATCo-R accounts for the performance of tasks such as visual monitoring, communication with aircraft crews, ATC coordination, and conflict detection and reaction. The model includes dynamic representations of the situation awareness about the aircraft and the alerts, a cognitive control mode and task scheduling. ATCo-R visually monitors the traffic situation on the runway and is supported the ATC alerts. ATCo-R may detect a safety-critical situation if AC-TX is observed to have passed the stopbar, or due to a stopbar violation alert, or due to a runway incursion alert. Following detection of the safety-critical situation, ATCo-R instructs both AC-TX and AC-TO to hold.

B. Risk assessment results

A key result of the Monte Carlo simulations is the probability of collision between the aircraft taxiing and taking-off. Since collision risks considered in air traffic are small, simulation speed-up by risk decomposition has been applied. Results presented earlier [17][18] indicate that a wrong intent situation awareness of the pilot flying of the taxiing aircraft is a condition with a strong effect on the accident risk. For the comparison with the risk results of the ET approach, we focus on the condition that the pilot flying of the taxiing aircraft intends to proceed on a normal taxiway (i.e. without being aware to be heading to the runway crossing). In this situation the pilot of the taxiing aircraft crosses the runway without contacting the runway controller, which is the condition considered in the event sequence-based risk assessment.
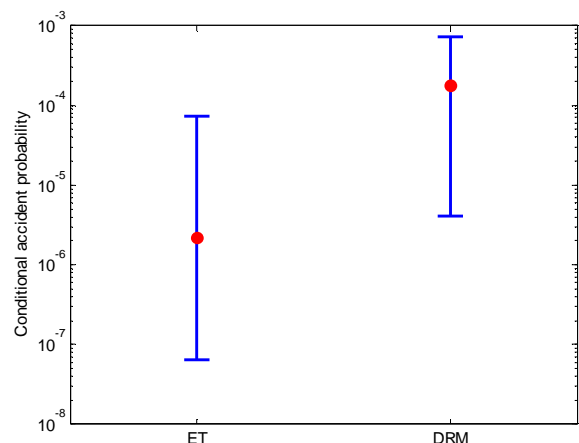

Figure 2: Conditional accident probability results of the ET study (lower/upper bound and geometric mean) and of the DRM study (95% uncertainty interval and point estimate).

To identify potential differences between model and reality and to evaluate their effect at the level of risk, a bias and uncertainty assessment method is an integrated part of the TOPAZ risk assessment methodology [20]. Results of a bias and uncertainty assessment are reported in [18] and they reveal that lack of knowledge on pilot performance contributes mostly to uncertainty in the risk. The point estimate and 95% uncertainty interval of the conditional accident probability given the runway incursion are shown in Figure 2. For comparison, it also shows the results achieved by the ET-based study.

V. EVENTS IN THE MC SIMULATIONS

A. Definition of events

To improve the insight in the performance of the agents in the DRM, the relation of this performance with the accident risk and to support the comparison with the event sequence-based analysis, we defined and recorded event occurrences in the Monte Carlo simulations of the agent-based DRM. As an onset for the analysis of event occurrences in the Monte Carlo simulations of the agent-based DRM, Figure 3 presents events for conflict recognition and collision avoidance actions by the agents as well as relations between these events. For instance, Figure 3 indicates that an active stopbar violation alert (event E8) may result in conflict detection by ATCo-R (event E5) and this event, on its turn, may result in warnings specified by ATCo-R towards the PFs of both aircraft (events E6 and E7). The times of first occurrence of most events were recorded in the MC simulations. The occurrence of events E1', E3' and E5' was inferred from the occurrence of related events.

B. Results of event occurrences

A total of 10 million Monte Carlo simulation runs were performed for the condition that the PF TX has the intent to proceed on a normal taxiway. In these runs a total of 1809 collisions were counted, which is consistent with the risk point estimate of 1.8E-4 for this condition found earlier. Table 2 shows the probabilities of the defined events and the conditional probabilities of these events given a collision. Key observations and explanations of the results in Table 2 are discussed next.

Pilot flying of the Taking-off Aircraft (PF-TO): PF-TO detects the conflict (event E1) in 99.2% of all simulated conflict scenarios. Here, PF-TO detects the conflict by own observation (event E1') in only 4.2% of all cases, whereas in the remaining 95.0% of all cases PF-TO detects the conflict via ATCo-R. Although PF-TO is very frequently monitoring the traffic situation and ATCo-R needs time to recognize the conflict and to warn PF-TO, the PF recognizes AC-TX as conflicting only if it is within a critical distance of 90 m to the runway centreline and ATCo-R can recognize AC-TX as conflicting as soon as it has passed the stopbar

Of the simulation runs ending in a collision, in hindsight we can see that PF-TO detects the conflict (event E1) in 99.7% of these cases, PF-TO detects the conflict by own observation (event E1') in 58.9% and via the controller in 40.8%. Thus for the conditional case given a collision it is found in hindsight that the probability of conflict detection by PF-TO is higher than in the unconditional case and the contribution of ATCo-R to detection of the conflict by PF-TO is significantly lower than in the unconditional case.
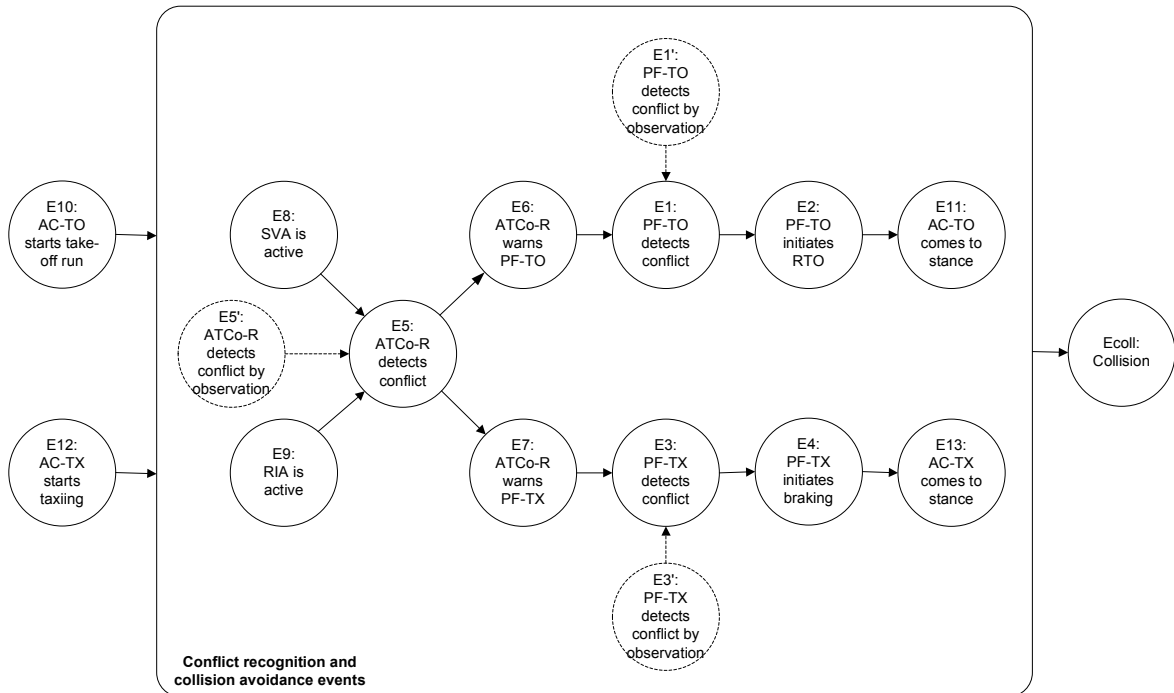

Figure 3: Relations between events in the MC simulations of the DRM. Events in solid circles are recorded in the MC simulations, events in dashed circles are inferred from the relative timing of recorded events.

Agent	Event		Probabilities	
	ID	Description	P(Eq)	P(Eq|Ecoll)
PF-TO	E1	Detects conflict	0.992	0.997
PF-TO	E1'	Detects conflict by own observation	0.042	0.589
PF-TO	E2	Initiates rejected take-off	0.566	0.239
PF-TX	E3	Detects conflict	0.998	0.913
PF-TX	E3'	Detects conflict by own observation	0.221	0.751
PF-TX	E4	Initiates braking	0.688	0.713
ATCo-R	E5	Detects conflict	0.993	0.999
ATCo-R	E5'	Detects conflict by own observation	0.393	0.228
ATCo-R	E6	Warns PF-TO	0.993	0.954
ATCo-R	E7	Warns PF-TX	0.993	0.569
ATC System	E8	Stopbar violation alert is active	0.940	0.999
ATC System	E9	Runway incursion alert is active	0.341	0.999
AC-TO	E10	Start take-off run	1	1
AC-TO	E11	Come to stance	0.566	0.000
AC-TX	E12	Start taxiing	1	1
AC-TX	E13	Come to stance	0.697	0.295
AC-TO AC-TX	Ecoll	Collision	1.81E-4	1

Pilot Flying of Taxiing Aircraft (PF-TX): PF-TX detects the conflict (event E3) in 99.8% of all simulated conflict scenarios. Here, PF-TX detects the conflict by own observation (event E3') in 22.1% of the cases, whereas in the remaining 77.7% of all cases PF-TX detects the conflict via ATCo-R. Although ATCo-R needs time to recognize the conflict and to warn PF-TX, the PF detects the conflict situation if it is recognized that AC-TO is taking off, whereas ATCo-R can already recognize the conflict as soon as the taxiing aircraft has passed the stopbar.

Of the simulation runs ending in a collision, in hindsight we can see that PF-TX detects the conflict (event E3) in 91.3% of these cases, PF-TX detects the conflict self (event E3') in 75.1% and via the controller in 16.2%. Thus for the conditional case given a collision it is found in hindsight that the probability of conflict detection by PF-TX is considerably lower than in the unconditional case and the contribution of ATCo-R to detection of the conflict by PF-TX is also significantly lower than in the unconditional case.

Runway Controller (ATCo-R): ATCo-R detects the conflict (event E5) in 99.3% of all simulated conflict scenarios. Here, ATCo-R detects the conflict by own observation (event E5') in 39.3% of all cases, whereas in the remaining 60.0% ATCo-R detects the conflict via the ATC alerting systems.

In the simulation runs ending in a collision, ATCo-R detects the conflict (event E5) in 99.9% of these cases. Here, ATCo-R detects the conflict by own observation (event E5') in 22.8% of these cases and via the ATC alert system in 77.1% of these cases. Thus for the conditional case given a collision it is found in hindsight that the probability of conflict detection by ATCo-R is considerably larger than in the unconditional case and the contribution of the ATC alert system to detection of the conflict by ATCo-R is somewhat higher than in the unconditional case.

The controller warns the pilots of the aircraft (events E6, E7) in 99.3% of all simulated conflict scenarios, which is equal to the detection rate by the controller (event E5). In the runs ending in a collision, the probability of a warning is decreased to 95.4% for PF-TO and to 56.9% for PF-TX. A factor contributing to the larger decrease for PF-TX is that in this conflict scenario, PF-TX is not on the R/T frequency of ATCo-R and their communication is thus delayed.

ATC Alerts: The stopbar violation alert is active (event E8) in 94.0% of all scenarios and in 99.9% of the cases ending in a collision. Mostly, it is not activated in situations that AC-TX stops close after the stopbar, such that the alert threshold has not yet been passed.

The runway incursion alert is active (event E9) in 34.1% of all scenarios and in 99.9% of the cases ending in a collision. It is not activated in situations where AC-TX taxies in front of AC-TO while it has not initiated take-off, or when AC-TX taxies after AC-TO has passed the crossing position.

Taking-off Aircraft (AC-TO): PF-TO initiates a rejected take-off (RTO) (event E2) in 56.6% of all cases and also in 56.6% of all cases AC-TO comes to stance (event E11). For the cases ending in a collision, an RTO was initiated in 23.9% of the cases and the aircraft came to stance in 0.0% of the cases.

Taxiing Aircraft (AC-TX): PF-TX initiates braking (event E4) in 68.8% of all cases and in 68.7% of all cases AC-TX comes to stance (event E13). In the cases that ended in a collision, braking was initiated in 71.3% of these cases and the aircraft came to stance in 29.5% of these cases.

Other detailed results (not shown) indicate that AC-TO is predominantly well within the first 500 m of the runway when the conflict is detected by either of the agents (PF-TO, PF-TX, ATCo-R, ATC System) or when the agents take action to prevent an accident. In contrast, for the cases ending in a collision, these events often occur when AC-TO is between 500 m and 1000 m; only for the detection of the conflict by the controller and the ATC alerts a considerable part of the PDF is below 500 m.

Similar results (not shown) for AC-TX show that overall the front-wheel of AC-TX is predominantly within 100 m from the runway centre-line when the conflict is detected by PF-TO or PF-TX, and when they start their collision avoiding actions. Overall, the controller detects the conflict at an earlier stage, predominantly when the AC-TX is between 150 and 100 m, and this range overlaps with that of the stopbar violation alert. However, at the time that the controller has warned the pilots, AC-TX is predominantly already within 100 m from the runway centre-line. There is a considerable overlap between the cores of the PDFs of the position of AC-TX in general and given the occurrence of a collision.

VI. SIMULATION OF CONDITIONS WITH AGENTS IN/OUT OF MONITORING ROLES OR CONTROL LOOPS

The results of the analysis in last section provided insight in the performance of the various agents in the runway incursion scenario and its relation with collision risk. To better understand the potential of agents to restrict the risk increase in cases where the performance of other agents is affected, we performed additional Monte Carlo simulations in which we placed one or more agents out of the monitoring role or control loop. This was done for all the agents that are capable of detecting a conflict, namely PF-TO, PF-TX, ATCo-R and ATC System. The conditions for placing these agents out of the monitoring role or control loop are:

- PF-TX does not actively monitor the traffic situation visually, such that PF-TX may only detect a conflict via a call of ATCo-R;

- PF-TO does not actively monitor the traffic situation visually, such that PF-TO may only detect a conflict via a call of ATCo-R;

- ATCo-R cannot communicate with the pilots;

- ATC System does not specify alerts.

These conditions for placing agents out of the monitoring role or control loop refer to the situation at the start and during the runway incursion scenario. These conditions were not assumed to hold prior to the occurrence of the runway incursion scenario.

For all relevant combinations of agents in or out of the monitoring or control loop, the conditional collision risk of the runway incursion scenario considered in this paper was determined by Monte Carlo simulation. This gives rise to 12 relevant combinations of conditions, which are shown in Table 3. Note that for conditions where ATCo-R is out of the control loop, it does not matter whether or not the ATC alerts are included in the control loop, as these can only be effective via ATCo-R. The runway incursion scenario considered earlier is case T1. For convenience Table 3 includes risk factors with respect to the lowest risk as obtained for case T1.

TABLE 3: CONDITIONAL COLLISION RISK RESULTS FOR VARIOUS CONDITIONS WITH AGENTS IN ('YES') OR OUT ('NO') OF THE MONITORING/CONTROL LOOP.

Case	PF-TX (monitor)	PF-TO (monitor)	ATCo-R (control)	Alerts (control)	Risk	Risk factor
T1	yes	yes	yes	yes	1.8E-4	1
T2	no	yes	yes	yes	1.0E-2	56.6
T3	yes	no	yes	yes	3.4E-4	1.89
T4	yes	yes	no	yes/no	2.2E-4	1.22
T5	no	no	yes	yes	1.7E-2	94.4
T6	no	yes	no	yes/no	1.7E-2	94.4
T7	yes	no	no	yes/no	1.9E-2	106
T8	no	no	no	yes/no	9.4E-2	522
T9	yes	yes	yes	no	1.9E-4	1.06
T10	no	yes	yes	no	1.2E-2	66.7
T11	yes	no	yes	no	2.1E-3	11.7
T12	no	no	yes	no	3.4E-2	189

Next we discuss key results of Table 3 in relation with the earlier presented results on events in the MC simulations of the runway incursion scenario (Table 2).

The collision risk of the runway incursion scenario increases by a factor 522 if none of the agents would be actively monitoring the traffic situation (case T8). In this case an accident is thus only prevented by chance, especially by the coincidental timing of the runway incursion with respect to the start of the take-off run. The accident risk of case T8 thus forms an upper bound for this particular runway incursion scenario.

The collision risk of the runway incursion scenario increases by a factor 1.06 if the ATC alert systems are not available (case T9). Stated differently, the presence of an ATC alert system barely reduces the collision risk. This is remarkable given the results for events E5 and E5' (Table 2), which show that if the ATC alert system is available, it warns ATCo-R before ATCo-R detected the conflict by own observation in 60% of the cases. Although the ATC alert system thus effectively supports ATCo-R, the results for case T9 show that the agents can well cope without the alerting system. In particular, even though the controller now regularly recognizes the conflict later, the conflict recognition time by the controller and by the pilots is only affected to a limited extent, such that the risk is increased by a factor 1.06 only.

The collision risk of the runway incursion scenario increases by a factor 1.22 if ATCo-R is out-of-the-loop (case T4). Thus the performance of ATCo-R in the resolution of the runway scenario has a small effect only on reducing the collision risk. This result may be seen as quite surprising, given the results for events E1, E1', E3 and E3' (Table 2) showing that the controller warns the pilots flying of the taking-off and taxiing aircraft in 95% and 78% of all cases before they have detected the conflict themselves. Notwithstanding this good performance of the controller, if the controller is placed out of the control loop in the modelled scenario, pilots can mostly detect the conflict themselves and react timely to avoid a collision, such that the risk increase is small.

The collision risk of the runway incursion scenario is increased by a factor 1.89 in the (hypothetical) case that PF-TO is not actively monitoring the traffic situation, but might still be warned by ATCo-R (case T3). If in addition to the lack of monitoring by PF-TO also ATCo-R is out of the control loop (case T7), then the risk is majorly higher by a factor 56 with respect to case T3. ATCo-R often warns PF-TO at an early stage, namely if AC-TO is well within the first 500 m of the runway. This early stage warning implies that ATCo-R can considerably restrict the risk increase of a non-monitoring PF-TO, as is manifest from the comparison of the risk factors in cases T3 and T7.

The collision risk of the runway incursion scenario is increased majorly by a factor 56.6 in the (hypothetical) case that PF-TX is not actively monitoring the traffic situation, but might still be warned by ATCo-R (case T2). If in addition to the lack of monitoring by PF-TX also ATCo-R is out of the control loop (case T6), then the risk increases by a factor 1.7 with respect to case T2. AC-TX is often close to the runway when ATCo-R warns PF-TX (event E7). Then warnings of

ATCo-R to PF-TX are often too late to prevent AC-TX entering a collision-critical area. Therefore, ATCo-R can barely restrict the risk increase due to a non-monitoring PF-TX, as is manifest from the comparison of the risk factors in cases T2 and T6.

The collision risk of the runway incursion scenario is increased majorly by a factor 94.4 in the case that only ATCo-R would be monitoring (while supported by the ATC alert system) and the pilots of both aircraft would not be monitoring, but may be warned by ATCo-R (case T5). The attained risk level is similar to the other cases where only one human operator is actively monitoring the traffic situation (cases T6 and T7). It shows that only one human actively monitoring human cannot effectively restrict the risk increase due to the malperformance of other operators.

Cases T10, T11, T12 represent situations where the ATC alert system is not available and also one or both of the pilots flying are not actively monitoring the traffic situation. It follows from comparison with the similar cases including the ATC alert system (i.e. cases T2, T3 and T5, respectively) that the effect of the non-availability of the ATC alert systems varies a lot.

- In the cases without active monitoring by PF-TX (T10 versus T2) the risk increases by a factor 1.2 only, indicating that the alerts are often too late to warn the PF-TX.

- In the cases without active monitoring by PF-TO (T11 versus T3) the risk increases by a factor 6, indicating that in this context the ATC alerts often warn ATCo-R such that ATCo-R can timely warn PF-TO.

- In the cases without monitoring by both pilots (T12 versus T5) a risk increase by a factor 2 is achieved, which is intermediate between the above indicated values.

These results indicate that the potential effectiveness of the ATC alert system can be better than the factor 1.06 found in case T9 if one or both pilots underperform. In the context given it is most important for timely warning of PF-TO.

VII. DISCUSSION

In this paper we compared risk assessment studies of a particular runway incursion scenario by an ET approach versus a multi-agent DRM approach. The focus in this paper is on a comparison of quantitative differences attained. Nevertheless, already at the qualitative level it can be argued that for the considered runway incursion scenario the ET-based risk model has clear limitations with regard to the representation of the dynamics of the scenario, the interactions between agents in the scenario, the variability of the performance of the agents in the scenario and the contextual conditions of the scenario. As a result of such limitations, the ET approach lacks transparency of the development of the risk model, the quantification of the event probabilities, the risk results and the feedback to the design. At a qualitative level it can be argued that the multi-agent DRM uses direct representations of the dynamics, agents' interactions, performance variability and contextual conditions, and as a result attains a better transparency for the development

of the risk model, the quantification of its parameters, the explanation of its results and the feedback to design.

A. Risk levels and contributions

Figure 2 shows that the accident risk was assessed to be considerably lower by the ET-based assessment in comparison with the DRM-based assessment. In particular the mean risk assessed by the ET is a factor 82 below the risk point estimate of the DRM.

The ET-based results for the risk reduction contributions of agents shown in Table 1 indicate that pilots reduce the risk by about a factor 2600, the controller reduces the risk by about a factor 2 and the ATC alert system supports a risk reduction by a factor 16. The DRM-based study shows that the level of risk is only manifest from the totality of the performance and interactions of all human operators and technical systems. As such, a overview of risk reduction contributions of different agents such as provided by the ET study cannot be derived by the DRM approach. Rather differences in risk between different constellations of agents being in or out of the monitoring role or control loop can be derived, such as shown in Table 3. It follows from this table that the risk is only reduced by a factor 1.06 by the ATC alert system. This is in contrast with the design objective of the ATC alert system to significantly reduce the runway incursion risk, as well as with the risk reduction factor 16 by the ATC alert system such as assessed by the ET approach.

B. Events in ET and DRM

In this study we showed a variety of events and their probabilities in the ET- and DRM-based safety assessments. With respect to the values of the event probabilities, a key difference between the approaches is that in the ET-based analysis they are mostly input, whereas in the DRM-based analysis they are output. In particular, in the ET-based assessment the event probabilities were based on interviews with operational experts, who expressed their opinion on the possibilities to recognize and resolve conflicts at a particular stage. Only for the incident and accident events the ET-based assessment provides probability values as output. In contrast, in the DRM-based assessment the probability values of the shown events are all outcomes emerging from the MC simulations of the DRM, whether they refer to events for conflict recognition and/or resolution by agents or to aircraft collisions. In particular we obtained the event probabilities by evaluating a large number of MC simulation runs of the runway incursion scenario, with the variability in the performance of the agents as specified in the DRM. The thus obtained event probabilities could be related to the occurrence of collisions and to variables of agents (e.g. aircraft positions), and a variety of event combinations could be evaluated. As such a considerably more diverse overview of relations between events and collision risk could be obtained by the DRM-based approach.

C. Safety analysis for feedback to design

Designers of an operation need to know main risk contributors and effective risk reduction means. Such risk analysis knowledge helps them to optimize the design from a

safety perspective. There are a number of methods for such risk analysis in the DRM-based approach:

A. Bias and uncertainty assessment of the DRM-based accident risk includes an evaluation of the sensitivity of the risk for changes in parameter values of the DRM. Operational aspects with large effects on the are focus points for designers [18].

B. The evaluation of the occurrence of conflict detection and resolution events of the agents, both unconditional and conditional given a collision, and the evaluation of performance variables (e.g. aircraft position) given the event occurrences (Section V).

C. The evaluation of risk effects due to placing agents out of monitoring or control loops. It gives insight in accident risk variations of the runway incursion scenario and in the capability of agents to compensate the lack of detection or control actions of other agents.

Methods A and C show that the accident risk is not sensitive for some aspects of the operation in the good visibility context considered. For instance, the risk would increase only by a factor 1.06 without an ATC alert system and it would increase only by a factor 1.22 if the controller would be out of the control loop at all. However, the risk is quite sensitive for some other aspects of the operation and the risk may increase by up to a factor 500 if none of the agents would be monitoring or in the control loop and a collision is only avoided by sheer luck.

Comparison of the results achieved by methods B and C clearly show that the risk is not manifest from the performance of individual human operators and technical systems, nor from the sole relations between human operators and/or technical systems, but only from the totality of the performance and interactions of all human operators and technical systems in the operational context considered. In particular, it follows from the analysis by method B that in about 94% of the runway incursion scenarios at least one of the alert types is active and in 60% of the scenarios the alert system warns the controller before (s)he has detected the conflict independently. Nevertheless, the analysis by method C shows that the risk increases only by a factor 1.06 without an ATC alert system. The reasoning is even stronger for the contribution of the controller. The model results indicate that the controller detects the conflict and warns the pilots in 99.3% of the cases and that in 95% and 78% of the cases the controller is able to warn the pilots flying of the taking-off or taxiing aircraft, respectively, before they have detected the conflict independently. In spite of this laudable performance of the controller in the model, the accident risk would only increase by a factor 1.22 if the controller would not play a role at all in the resolution of the runway incursion scenario. It is only by considering the totality of the interactions between the agents and the variability in their performance in huge numbers of simulations that reveals the effects on accident risk due to aspects of the operation.

Insight in effective impact on the risk of taking agents out of the monitoring role or control loop was not obtained by the ET approach. For instance, the ET-based results suggest that the risk is reduced by a factor 16 due to the ATC alert system.

In this ET a change in the operation, such as leaving out an ATC alert system, would imply that alert-related events cannot occur. Assuming that the other event probabilities remain the same, this would lead to a major risk increase by a factor 16. However, the assumption that the other event probabilities remain the same appears not to be true. In conclusion, the ET based analysis does not effectively support safety analysis for feedback to design.

D. Implications for expert judgement

As a result of the conclusion that the level of safety need not be manifest from the performance of individual human operators and technical systems, nor from the sole relations between human operators and/or technical systems, it also follows that assessing the contributions for prevailing accidents by interviewing single operators (pilots and controllers) and by judging their contributions, does not well account for the complexity of the interactions in conflict scenarios and thereby may well lead to inaccurate safety assessment results. For instance, based on controller interviews it was assessed in the ET-based study that the controller, when supported by an ATC alert system, would have a large effect on reducing the accident risk of the runway incursion scenario. However, for an individual controller it is not well possible to judge the probability that a controller warning reaches the pilots before they have detected the conflict independently. Even more importantly, it is not possible for the individual controller to quantify the effectiveness of a controller warning at the level of accident risk reduction, since it supposes an evaluation of all other possibilities of other agents to detect and resolve the conflict scenario.

E. Implications for human-in-the-loop simulations

The contrast between the seemingly good performance of a human operator and the limited effect of this performance on the accident risk in a conflict scenario, as found by the large scale Monte Carlo simulations of the multi-agent DRM-based study in this paper, poses limitations on the safety conclusions that can be attained by other types of simulations. In the air traffic control domain, new concepts are regularly evaluated by human-in-the-loop simulations, in which the performance of (real) air traffic controllers is evaluated in a simulated environment. For operations on the airport this is done in tower simulators, where simulated aircraft movements on the aerodrome are projected in a 360 degrees view, the controllers are supported by their usual ATC systems (which may include alerts) and the controllers can communicate with pseudo-pilots who control the movements of the simulated aircraft. The numbers of aircraft handled in such simulations are similar to what can be achieved in reality, e.g. a runway controller may handle about 25 to 40 aircraft per hour. Human-in-the-loop simulation experiments typically last several days and often aim to evaluate several configurations, typically leading to some hundreds of aircraft handled in a particular configuration. In human-in-the-loop simulations occasionally conflict scenarios may be instantiated and the effectiveness of a controller to detect the conflict and warn pilots may be evaluated. Whereas it is manifest that the numbers of conflict scenarios that can be evaluated in human-in-the-loop simulations are far too small to evaluate safety up to the level

of accident risk, the results of this paper moreover indicate that results on the performance of human operators in such simulations say little about their contributions to safety. Consider, for instance, a hypothetical result of a human-in-the-loop simulation experiment that a controller is able to warn the pilots in conflict situations in the large majority of conflicts (say 95%). This might be interpreted as an indication that the controller is contributing considerably to avoiding accidents, thus forming an important safety barrier. However, the presented results provide an example where the controller warns the pilots in 99% of the cases and still the accident risk would increase only slightly without any contributions of the controller due to the performance of the other agents in the operation. More in general, the results of this paper indicate that if the number of simulations is not sufficient to estimate the accident risk of a conflict scenario, it is hard to judge from the performance of individual agents what their effect on safety at the level of accident risk may be.

As a way forward for using human-in-the-loop simulations in safety assessment, aspects of the performance variability of human operators in safety relevant scenarios may be measured and such measurement results may be used to support the development of appropriate agent models in a DRM. Detailed discussion of such coupling of human-in-the-loop simulations and dynamic risk modelling is out of the scope of this paper.

VIII. CONCLUSION

Considerably different results were obtained in the accident risk assessments of the complex socio-technical system involved in the runway incursion scenario. The Monte Carlo simulations of the multi-agent DRM uniquely show that the risk is not manifest from the performance of individual human operators and technical systems, nor from the sole relations between human operators and/or technical systems, but only from the totality of the performance and interactions of all human operators and technical systems in the operational context considered. These findings imply that judging the contributions of single human operators or technical systems for prevailing accidents may neglect the complexity of the interactions in socio-technical systems and thereby lead to inaccurate safety assessment results.

In conclusion, in this paper we showed that not only in theory but also in actual safety assessment of a realistic air traffic operation, multi-agent dynamic risk modelling has a considerable number of advantages over event sequence-based approaches. We have also shown that the findings have significant ramifications for the evaluation and testing of novel operations in air traffic management: commonly applied analysis processes, such as human-in-the-loop simulations, model development, model validation and feedback to design, appear to have a serious lack in capturing the safety related impacts of interactions between the multiple agents involved in such novel operations.

REFERENCES

[1] Turner BA. Man-made disasters. Wykeham Science Press, London, UK, 1978

[2] Perrow C. Normal accidents: Living with high-risk technologies. Basic Books, New York, USA, 1984

[3] Hollnagel E. Barriers and accident prevention. Ashgate, Aldershot, England, 2004

[4] Eurocontrol. Air navigation system safety assessment methodology. SAF.ET1.ST03.1000-MAN-01, edition 2.0, 2004

[5] EUROCAE. ED78A Guidelines for approval of the provision and use of ATS supported by data communication, 2000

[6] Leveson N. A new accident model for engineering safer systems. Safety Science 42:237-270, 2004

[7] Blom HAP, Bakker GJ, Blanker PJG, Daams J, Everdij MHC, Klompstra MB. Accident risk assessment for advanced air traffic management. In: Donohue GL and Zellweger AG (eds.), Air Transport Systems Engineering, AIAA, pp. 463-480, 2001

[8] Bedau MA. Weak emergence. In Tomberlin J (ed.), Philosophical Perspectives: Mind, Causation, and World, Vol 11, pp. 375-399, Blackwell, Malden (MA), USA, 1997

[9] Corning PA. The re-emergence of "emergence": A venerable concept in search of a theory. Complexity 7(6):18-30, 2002

[10] Dulac N, Owens B, Leveson N et al. Demonstration of a new dynamic approach to risk analysis for NASA's constellation program. MIT SCRL Report, March 2007

[11] Blom HAP, Stroeve SH, Scholte JJ, De Jong HH. Accident risk analysis benchmarking Monte Carlo simulation versus event sequences. Proceedings Third International Conference on Research in Air Transportation (ICRAT 2008), Fairfax (VA), USA, 1-4 June 2008

[12] Stroeve SH, Blom HAP, De Jong HH, Scholte JJ. Comparison of accident risk assessment by event sequence analysis versus Monte Carlo simulation. Eurocontrol Safety R&D Seminar, Southampton, UK, 22-24 October 2008

[13] ICAO. Manual on the prevention of runway incursions. International Civil Aviation Organization, Doc 9870 AN/463, first edition, 2007

[14] Eurocontrol. European action plan for the prevention of runway incursions, release 1.1, 2004

[15] Scholte JJ, Blom HAP, Van den Bos JC, Jansen RBHJ. Management of ATM performance in operational concept development and validation: a case study. Proceedings of Eight USA/Europe ATM R&D Seminar, Napa, USA, 2009

[16] De Jong HH, Tump RS, Blom HAP, Van Doorn BA, Karwal AK, Bloem EA. Qualitative safety risk assessment of a RIASS based operation at Schiphol airport including a quantitative model: Crossing of departures on 10L/19R under good visibility conditions. National Aerospace Laboratory NLR, memorandum LL-2001-017, Amsterdam, The Netherlands, 2001

[17] Stroeve SH, Blom HAP, Van der Park MNJ. Multi-agent situation awareness error evolution in accident risk modelling. Proceedings 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary, 2003

[18] Stroeve SH, Blom HAP, Bakker GJ. Systemic accident risk assessment in air traffic by Monte Carlo simulation. Safety Science 47:238-249, 2009

[19] Everdij MHC, Klompstra MB, Blom HAP, Klein Obbink B. Compositional specification of a multi-agent system by stochastically and dynamically coloured Petri nets. In: Blom HAP, Lygeros J (eds.), Stochastic Hybrid Systems, LNCIS 337, Springer-Verlag, pp. 325-350, 2006

[20] Everdij MHC, Blom HAP, Stroeve SH. Structured assessment of bias and uncertainty in Monte Carlo simulated accident risk. Proceedings of the 8th International Conference on Probabilistic Safety Assessment and Management, May 14-18 2006, New Orleans, USA, 2006