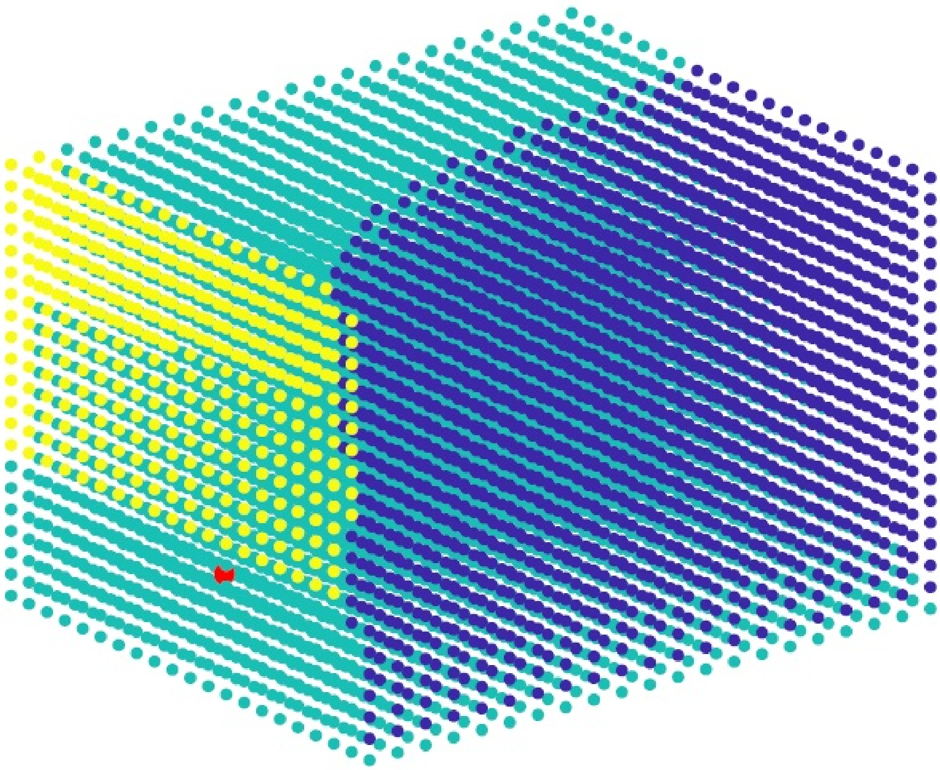
Multi-criteria decision making for improvement of security and efficiency at airport security checkpoints using agent-based models

Adin Mekic

Technische Universiteit Delft



Multi-criteria decision making for improvement of security and efficiency at airport security checkpoints using agent-based models

by

Adin Mekic

in partial fulfillment of the requirements for the degree of

Master of Science
in Aerospace Engineering

at the Delft University of Technology.

Student number:	4303687	
Project duration:	February 25, 2019 – December 4, 2019	
Thesis committee:	Prof. dr. R. Curran,	TU Delft, chair
	Dr. O. A. Sharpanskykh,	TU Delft, supervisor
	S.A.M. Janssen Msc,	TU Delft, daily supervisor
	Dr. ir. D. Pool,	TU Delft, external board member

Acronyms

AATOM An Agent-based Airport Terminal Operations Model.

AHP Analytic Hierarchy Process.

BWM Best Worst Method.

DCA Discrete Choice Analysis.

DCE Discrete Choice Experiment.

ELECTRE ELimination Et Choix Traduisant la REalité.

FPR False Positive Rate.

FUCOM FUll COnsistency Method.

MADM Multi Attribute Decision Making.

MAUT Multi Attribute Utility Theory.

MAVT Multi Attribute Value Theory.

MCAP Multi Criteria Aggregation Procedure.

MCDM Multi Criteria Decision Making.

MODM Multi Objective Decision Making.

PROMETHEE Preference Ranking Organization METHod for Enrichment of Evaluations.

SLAM Simple Landside Aggregate Model.

SMART Simple Multi-Attribute Rating Technique.

TOPSIS Technique for Order of Preference by Similarity to Ideal Solution.

TPR True Positive Rate.

VIKOR VIsekriterijumska Optimizacija I KOmpromisno Resenje.

List of Figures

List of Tables

CONTENTS

I

Paper

Multi-criteria decision making for improvement of security and efficiency at airport security checkpoints using agent-based models

Adin Mekic, Stef Janssen, Alexei Sharpanskykh

***Abstract*—Agent-based modelling and simulation has shown to be a suitable paradigm for analyzing complex airport systems, such as airport security checkpoints. The models can be used to analyze important airport terminal performance areas, such as security and efficiency. However, limited knowledge exists on how agent-based models can be used to aid airport managers in their decision making. The agent-based models allow to estimate the criteria of the alternatives of a decision problem, but without criteria weights and score aggregation, the alternatives can not be compared or ranked. The main branch of techniques to establish weights for multiple criteria such that a score can be aggregated for decisions is called multi-criteria decision making. Thus, integrating multi-criteria decision making methods with agent-based models is a method to provide decision support for airport managers. Therefore, this research aims to answer how multi-criteria decision making methods can be integrated with agent-based models to provide decision aid to airport terminal decision makers with the aim of improving security and efficiency. To answer this research question, a methodology was proposed which combines agent-based models, multi-criteria decision making, discrete choice models, and expert knowledge. The methodology is applied to a case study in which several alternatives of configuring personnel are available. The results show that efficiency focused operators, which have low accuracy, introduce most variance between best and worst alternatives. The best alternatives have efficiency focused operators placed on peak hours because they decrease the queue in the peak hours, while maximizing overall hit rate because less passengers pass per lane in peak hours compared to non-peak hours. For realistic variations of operator performance, operator placement is found to have a significant influence on security checkpoint performance.**

***Index Terms*—Agent-based simulation, Multi-criteria decision making, Discrete Choice Modelling, Airport Security, Airport Efficiency, Trade-off**

I. Introduction

The never-ending growth of aviation poses significant challenges towards the operation of airports. This stimulates the need for methods that improve complex decision making with regard to airport operations. Airport decision makers aim to maximize the utility of their available resources. Often, multiple conflicting criteria are involved in these decision making problems such as security, safety, efficiency, cost, and passenger satisfaction. The classic operations research method of approaching a decision problem is to define a single objective function and optimize it for a set of actions. This is a reductive and unnatural way to look at this type of decision problems in which inherent conflicts are present [1].

Two important airport performance areas to improve and optimize for are security and efficiency. Many studies analyze security and efficiency separately, but few studies explore the relationships between them. Several exceptions exist. Wilson et al. utilizes simulations which allows a wide range of analysis such as policies, security effectiveness, resource utilization, and cost, but lacks a formal methodology [2]. Janssen et al. use agent-based models for security risk assessment and efficiency estimation, to show that security and efficiency are not always conflicting objectives [3]. Knol et al. use agent-based models to show the impact of varying security focus, a variable which increases the True Positive Rate (TPR) and False Positive Rate (FPR) of the system if increased, and decreases the TPR and FPR if decreased [4]. Overall, practical applications of these simulation models remain limited because they are not directly interpretable by airport managers.

Several factors are of importance regarding security effectiveness. Yoo defined human resources, facility/equipment, and procedures/responsibility, as major factors affecting security effectiveness [5]. He analyzed their importance concluding that human resources are the main factor. It is an often recurring theme that human factors play a critical role in improving security systems [5]–[9]. Skorupski found that the efficiency of cabin baggage screening varies significantly at different security checkpoints within the same airport, and attributed this to human factors. A human element that is partially responsible for this is informal behaviour where security decision are made despite rules and regulations [6], [10]. Another relevant human element is performance. Operator performance depends on aptitude abilities, visual knowledge, motivation, and correction action [11], leading to a different performance for each operator. These type of operator characteristics require a microscopic scale model. Agent-based models are often used microscopic scale models to model actions and interaction of agents. Differences in performance of security operators have not been analyzed through simulation studies, in terms of multiple criteria. This is an identified research gap, which has to be addressed due to the large importance of human factors.

Agent-based modelling and simulation is a suitable paradigm to model the dynamic complex socio-technical airport terminal environment since it includes complex autonomous agents where many components act and interact with the environment. Multi-agent system features such as autonomy, local views, and decentralization make agent-based modelling suitable to model the operators and passengers in the airport terminal. The microscopic nature of agent-based models allows detailed specification of agents such that differences in human performance can be implemented and analyzed. However, the high complexity of agent-based models makes it difficult for industry experts to benefit from these models. Therefore, this research aims to integrate agent-based models with multi-criteria decision making to make decision aid using agent-based models more accessible. This integration does not receive wide attention in scientific literature, but exceptions exist such as [12]–[14]. Often, no formal methodology is presented [12], or the methodology applies to a very specific problem [13]. The work of Amato et al. does provide a general methodology, but follows a theoretical approach which does not provide details on the methodological steps. Therefore, this work presents an analytical approach to the methodology to better cover the different steps of the integration.

This paper addresses the following research question: how can multi-criteria decision making methods be integrated with agent-based models to aid decision makers in obtaining recommendations of configuring personnel in airport terminal operations with the aim of improving security and efficiency? Examples of recommendations are to follow a certain scheduling strategy, or indication of a best alternative. Agent-based models are used to simulate a security checkpoint in the airport terminal environment, such that security and efficiency performance can be measured and relationships between them can be determined. Multi-criteria decision making was used to evaluate different personnel configurations. Discrete choice modelling was used to incorporate expert knowledge into the multi-criteria decision making weighting method.

The paper is structured as follows. Section II discusses related work. Section III describes the system under investigation and its operational context. Section IV describes the methodology. Section V provides the experiments & results. Section VI presents the discussion. Section VII presents the conclusion.

II. Related Work

The agent-based model used in this research is an existing airport security checkpoint agent-based model [4], which was built on the AATOM architecture [15]. The Ratcliff diffusion model is incorporated with the agent-based model to provide the operators a more cognitive human way of decision making [4], [16]. McCauley's biomathetmatical fatigue model has been used to account for effects of sleep and sleep loss on operator performance [17]. Walsh has combined McCauley's fatigue model with Ratcliff's diffusion model to relate fatigue with decision making. More details on the model are found in Appendix B.

A sampling strategy is required to estimate agent-based model output parameters. Notable methods are simple random (Monte Carlo) sampling and latin hypercube sampling [18]. If the case study contains many alternatives, for example a scheduling problem, it will be required to create an surrogate model to be able to evaluate all alternatives. Machine learning can, for example, be used to build surrogates from agent-based models [19].

This paper specifically addresses Multi-Criteria Decision Making (MCDM) methods for a discrete solution domain, known as multi-attribute decision making methods. The objective of a MCDM method is to improve the quality of the decision. Many different classical MCDM methods exist which can be used such as AHP [20], PROMETHEE [21], ELECTRE [22], MAUT [23], VIKOR [24], TOPSIS [25], BWM [26], and FUCOM [27]. A review of their procedures, advantages, and disadvantages is found in Appendix A. Other non-classical approaches like verbal decision analysis, and multiobjective programming based methods for continuous solution domains exist [1].

The existence of many different MCDM methods requires to compare methods with respect to weight estimation, decision problematic, information type, procedure of aggregation, and result. These analyses and comparisons have been performed by many [28]–[30]. Even though the methods can be very different, they share the same high-level concept of establishing weights for criteria, and calculating scores for alternatives.

The main difference between different MCDM methods is the way in which the weights are obtained. Direct rating (e.g. WSM, TOPSIS), tradeoffs (e.g. MAUT), and pairwise comparison (AHP, ELECTRE, PROMETHEE), are existing ways in which preferences can be modelled. Tradeoffs and pairwise comparison methods often require experts to value criteria and alternatives which leads to inaccurate responses since evidence from cognitive psychology indicates that people cannot accurately report on why they make certain decisions [31].

Verbal decision analysis, an alternative method of using expert preference information for multi-criteria decision making based on cognitive psychology, uses ordinal tradeoffs with a psychological basis to address inconsistencies in human judgement. However, verbal decision analysis is developed for unstructured problems with mostly qualitative parameters with no model for aggregation [32]. Like verbal decision analysis, discrete choice analysis has a cognitive psychological basis, and provides a way of modelling expert information to obtain relative weights of criteria. It is also suitable for structured problems with quantitative parameters. It provides a different way of obtaining criteria weights.

If criteria have different units, it is required to perform normalization of data to be able to aggregate a single score. Five commonly used normalization techniques are linear max, linear max-min, linear sum, vector normalization, and logarithmic normalization [33].

III. Case Study

The environment of the case is an airport security checkpoint and its area where the objects of the model are a physical queue and a security lane, seen in Figure 1. The security checkpoint was chosen as the main point of analysis because it is the main operation for reducing forbidden items entering the aircraft. Thus, it is important for reducing security risks.

Two type of agents are defined in the model: operators and passengers. The goal of the passengers is to pass through the security checkpoint as quickly as possible. Passengers have carry-on luggage which can contain a forbidden item. The goal of the operators is to confiscate the forbidden item.

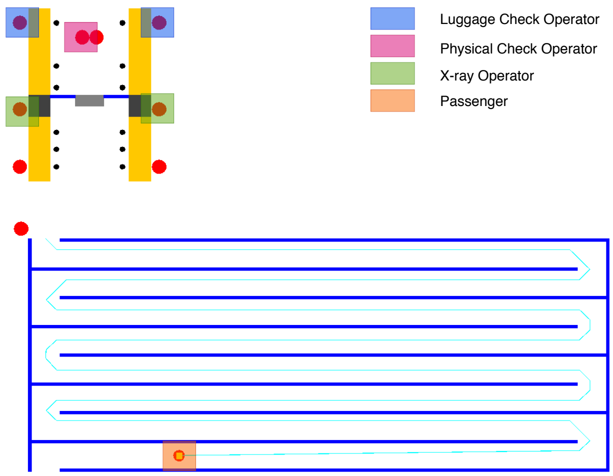


Fig. 1. The environment of the agent-based model.

Different security operator types are defined: x-ray, luggage check, and physical check, as seen in Figure 1. Each operator makes a decision on whether a forbidden item is present or not. For each lane a single X-ray operator and luggage check operator is present. For every two lanes, a single physical check operator is present. The performance of operators is distinguished based on response time (speed), true positive rate (accuracy), and hours of sleep (fatigue). Average queue time is selected as an efficiency performance indicator. The ratio of the number of confiscated forbidden items divided by the total number of forbidden items presented at the security check is used as a security performance indicator. Average operator cost is selected as a cost performance indicator.

The case study includes assigning airport security checkpoint personnel to slots. Since performance quality of employees varies, different personnel assignments will have a different impact on the performance indicators. The performance indicators are average queue time, hit rate, and average operator cost. The main operator schedule slots of the regional airport under investigation are a morning slot from 05:00 to 13:00, and an afternoon slot from 13:00 to 19:00. In general, for the airport under investigation, it was seen that there is often a peak of flights departing between 07:00 and 07:30. Table I shows the main flights of a chosen flight day. Flights f1-f6 are the main flights for the morning slot on that day in which four lanes are open. Flights f7-f9 are the main flights for the afternoon slot on that day in which two lanes are open.

TABLE I
SCHEDULED DEPARTING FLIGHTS USED IN THE AGENT-BASED MODEL.

Flight	Nr. of passengers	Departure time
f1	70	07:05
f2	130	07:05
f3	150	07:10
f4	140	07:25
f5	130	07:30
f6	135	07:35
f7	140	15:50
f8	140	16:25
f9	150	16:45

The two slots require a total of 15 operators following operator placement as seen in Figure 1. The four lanes in the morning slot require 4 O_{xr}, 4 O_{lc}, 2 O_{pc}, and the two lanes in the afternoon slot require 2 O_{xr}, 2 O_{lc}, and 2 O_{pc}. The assumption is made that an operator performs a single activity during the slot. Thus, 6 O_{xr} can be placed in 15 combinations across the two schedules. Likewise, 15 combinations exist for luggage check operators, and 3 combinations exist for the physical check operators. Thus, a total of 675 scheduling alternatives exist.

IV. Methodology and its Application

Figure 2 shows an overview of the methodology. These seven steps were taken to answer the research question. The ideas behind the methodology steps are as follows. Step 1 alters the used agent-based model such that the different scheduling alternatives of the case study can be tested. Step 2 calibrates the agent-based model such that the model represents reality as accurate as possible. Step 3 defines the criteria weights by using discrete choice experiment data collected from experts. Step 4 constructs an aggregation procedure such that the scheduling alternatives can be compared. Step 5 builds a surrogate of the agent-based model to reduce computational efforts. Step 6 analyzes the alternatives by finding the difference between the best and worst performing schedule. Step 7 provides decision aid. Steps 1-5 are addressed in this section. Step 6 and step 7 are described in Section V.

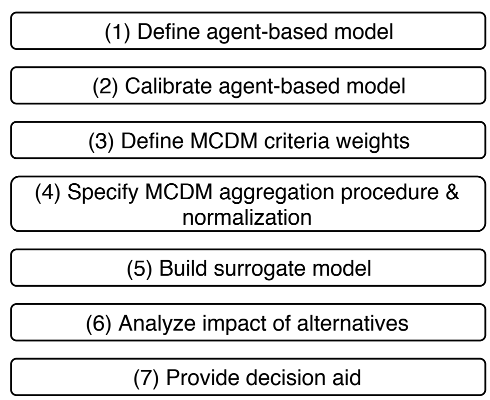


Fig. 2. Methodology Overview.

A. Step 1: Agent-based model definition

As mentioned in Section II, this study used an existing airport terminal agent-based model. Due to the large importance of human factors, the model was modified to incorporate parameters of individual operator performance. Operators are distinguished through speed, accuracy shift, accuracy and fatigue. Speed is γ, accuracy shift is α, accuracy is σ and ϕ is fatigue. $O_{xr}(\gamma,\alpha,\sigma,\phi)$, $O_{lc}(\gamma,\alpha,\sigma,\phi)$, and $O_{pc}(\gamma,\alpha,\sigma,\phi)$, denote the specific variables for an X-ray operator, luggage check operator, and physical check operator respectively.

More information on the model and the model modifications is found in Appendix B. Furthermore, a sensitivity analysis was conducted to better understand the behaviour and impact of the introduced parameters. This information is found in Appendix E.

B. Step 2: Agent-based model calibration

The model was calibrated such that it represents reality as accurately as possible. An existing calibration of the diffusion model, calibrated to experimental data of response times, was used as a baseline configuration [4]. The version of AATOM was upgraded as compared to AATOM version used in the baseline configuration. The baseline calibration was recalibrated to adjust for a realistic throughput.

The baseline calibration to a throughput of 2.6 in rush hour was not replicable. Empirically collected evidence suggests the throughput to be closer to 2.0. In the baseline calibration, a correction factor was applied to scale down the non-decision times, since the model is unable to reproduce a realistic security throughput rate with realistic processing times [34]. Besides the existing factor for non-decision time in the baseline calibration, correction factors for the decision time, luggage drop time, and luggage collect time, were introduced to calibrate the throughput. The values of the baseline calibration and the recalibration are seen in Table II.

TABLE II
CORRECTION FACTOR CALIBRATION VALUES FOR THE BASELINE MODEL AND THE RECALIBRATED MODEL.

Correction factor parameter	Baseline calibration	Recalibration
Non-Decision time	0.65	0.60
Decision time	1.0	0.50
Luggage drop time	1.0	0.50
Luggage collect time	1.0	0.50

Here the response time is equal to the sum of decision time and non-decision time. The decision time is obtained from the diffusion process, whereas non-decision time is a constant used for calibrating to accurate operators response times. The correction factors simply scale the values of the parameters. The correction factor of the decision time was decreased since the baseline calibrated model decision time was too large. This baseline decision time for X-ray operators was found to be $8.4s$. Since average response times of X-ray operators is found to be closer to $6s$ [35], and this $6s$ can be reduced to as low as $3.5s$ through adaptive computer-based training [8]. Therefore, the decision time correction factor was set to 0.5 such that sufficiently low X-ray operator responses can be explore in the agent-based model.

C. Step 3: MCDM criteria weighting

As discussed in Section II, MCDM methods often require experts to value criteria directly or value pairwise comparisons, which leads to inaccuracies since people cannot accurately report on why they make certain decisions [31]. It is observed that people cannot directly observe their cognitive process, but are able to make good choices between alternatives [31]. Therefore, discrete choice analysis was selected for obtaining criteria weights as it is a reliable way of obtaining weights.

To obtain information on weights, data from experts was collected through discrete choice experiments (DCEs). The attribute levels of the criteria were determined by a combination of agent-based model tests, DCE pilot tests, and security expert cooperation. Efficient designs based on the utility maximization decision rule were chosen as the experimental design for the choice sets. The design of the choice sets and the DCE was performed using a software tool [36]. Details on the DCE design theory, the DCE design, and how the results are obtained, are found in Appendix D.

Table III shows the criteria, their regression coefficients, the standard errors, the p-values, and their weights, which were obtained from the DCE. Ratios of regression coefficients are marginal rates of substitution. This is the rate at which one attribute can be given up for another while maintaining the same utility. Thus, $\frac{\beta_2}{\beta_1} = \frac{0.2989}{-0.1388} = -2.15$, implies that a hit rate unit was found to be the equivalent of 2.15 efficiency units, which means that 0.01 hit rate is valued 129 seconds of queuing time. The low value for cost, and its statistical insignificance, indicates that cost is not an influential parameter for trade-off.

TABLE III
THE REGRESSION COEFFICIENTS, STANDARD ERROR, P-VALUE, AND WEIGHTS ESTIMATED BY APPLYING MULTINOMIAL LOGIT REGRESSION ON THE DCE DATA.

Variable	Coefficient	Standard error	p-value	Weight
1: Efficiency	-0.139	0.046	0.002	0.296
2: Security	0.299	0.058	<0.001	0.637
3: Cost	-0.031	0.037	0.389	0.067

D. Step 4: MCDM aggregation procedure

The WSM method was used for the aggregation procedure as it provides a simple quantitative aggregation of a score without bias. The multi-criteria decision making problem can then be formulated as seen in Equation 1.

$$A = \begin{array}{c c} & \begin{array}{cccc} c_1 & c_2 & \dots & c_j \end{array} \\ \begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_i \end{array} & \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1j} \\ q_{21} & q_{22} & \dots & q_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ q_{i1} & q_{i2} & \dots & q_{ij} \end{pmatrix} \end{array} \tag{1}$$

Here a_i represents an alternative, c_j represents a criterion, and q_{ij} represents the score of alternative i with respect to criterion j. The overall score of an alternative, V_i, is the sum of all scores with respect to its criteria. Since the criteria have varying importance, they are assigned weights, w_j, where $\sum w_j = 1$ and $w_j \geq 0$. Thus, an overall score can be represented as Equation 2.

$$V_i = \sum_{n=1}^{n} w_j q_{ij} \tag{2}$$

Since the criteria have different units, it is required to perform normalization of data to be able to aggregate a single score, and rank the alternatives. Linear max-min is a suitable normalization method for this problem since it does not allow bias towards ideal solutions. This is desirable such that the aggregation in both directions occurs in equal magnitude.

Thus, if the relative weight of efficiency with respect to security is such that 200s:0.01, then the alternatives in Table IV should have the exact same score.

TABLE IV
EXAMPLE OF MCDM INPUT DATA.

	Efficiency	Security	Cost
Alternative 1	500 s	0.08	€ 14
Alternative 2	700 s	0.07	€ 14
Alternative 3	900 s	0.06	€ 14

A property from linear max-min that was undesired, is that if queue time of alternative 3 of example Table IV is increased, alternative 1 & alternative 2 are not valued equally. Therefore, the linear max-min equation for cost criteria was changed to Equation 3, such that it accounts for the weights using units. q_{ij} indicates the score of alternative i with respect to criterion j, u_j is the unit of measurement, r_{ij} is the value, and r_{min} the minimum value of an alternative i for criterion j. The unit of measurement for efficiency, security, and cost, are $u_1 = 60$, $u_2 = 0.01$, and $u_3 = 1$ respectively.

$$q_{ij} = \frac{r_{max} - r_{ij}}{r_{max} - r_{min}} \rightarrow q_{ij} = \frac{r_{max} - r_{ij}}{u_j} \tag{3}$$

Based on Equation 3, each alternative receives a score for each criterion if it performs better than the alternative that performs worst for that criterion. Thus, considering Table IV, alternative 1 will only receive score for efficiency, alternative 2 will receive score for both efficiency and security, and alternative 3 will receive only score for security. The difference between two aggregated scores indicates how much one alternative is better than the other in terms of the $\frac{u_j}{w_j}$. For example, consider two alternatives that perform equal on security and cost but differ on efficiency, if the difference in score between the two alternatives is 1.0, and $u_1 = 60$ & $w_1 = 0.3$, then the alternative with the higher score performs $200s$ better on efficiency. The introduction of u_j in the normalization gives meaning to the aggregated score. More information on the normalization procedure that is used is found in Appendix C.

E. Step 5: Surrogate model

Sampling each alternative n times exceeds the computational cost for scheduling problems, as it would take in the order of months to simulate these alternatives on a desktop pc. Thus, a surrogate model was built. A surrogate model is a substitute model which aims to replicate the behaviour of the original model, while using a fractional amount of the computational effort. The assumption is made that higher-order effects have a negligible impact on outcome. The assumption was made following an experiment that varies $O_{xr}(\gamma)$, $O_{lc}(\gamma)$, and $O_{pc}(\gamma)$ simultaneously to study the higher-order effect on the output, and an experiment which does this to $O_{xr}(\alpha)$, $O_{lc}(\alpha)$, and $O_{pc}(\alpha)$. The results indicated that an additive approximation of the output was within an error margin of 5%. However, more work is required to assess the actual higher-order effects that are present within the model since the experiments are two specific cases within high-dimensional solution space. The assumption allows to build a surrogate model with additive properties. For each operator, the first-order effects of speed, accuracy, and fatigue are estimated. The discretization scheme is seen in Table V, and contains 5 points for speed, 8 points for accuracy, and 3 points for fatigue. Discretization [4:2:8] implies from 4 to 8 with step size 2, thus [4 6 8] are the sampled points. The third column indicates what a value of the parameter translates to in terms of response time or TPR. Each point is sampled randomly 500 times for each slot. While variables are varied, other average operators remain constant at $O(\gamma, \alpha, \sigma, \phi) = O(1.0, -0.15, 0.95, 8)$, where average TPR was taken from empirical conducted studies [4], [37]. The graphs of the results which are used for the surrogate are found in Appendix F.

Based on these results, some simplifications were made in the surrogate model. For example, the impact of fatigue on queue time was a maximum of 2 seconds, thus was considered negligible. Furthermore, the overall hit rate is averaged across all passengers, whereas for queue time only the average queue time of the morning slot is used. The reason being that the average queue time of the afternoon slot was between 2 to 3 minutes for multiple experiments, which is considered a very acceptable average queue time.

TABLE V
DISCRETIZATION SCHEME OF SIMULATED PARAMETERS.

Variable	Discretization	Response time / TPR
$O_{xr}(\gamma)$	[0:0.4:1.6]	[4.50 5.97 7.44 8.90 10.37]
$O_{lc}(\gamma)$	[0.6:0.2:1.4]	[41.58 53.94 66.30 78.66 91.02]
$O_{pc}(\gamma)$	[0.6:0.2:1.4]	[24.12 30.66 37.20 43.74 50.28]
$O_{xr}(\sigma)$	[0.15:0.1:0.85]	[0.29 0.57 0.75 0.82 0.89 0.92 0.95 0.96]
$O_{lc}(\sigma)$	[0.15:0.1:0.85]	[0.29 0.57 0.75 0.82 0.89 0.92 0.95 0.96]
$O_{pc}(\sigma)$	[0.15:0.1:0.85]	[0.29 0.57 0.75 0.82 0.89 0.92 0.95 0.96]
$O_{xr}(\phi)$	[4:2:8]	-
$O_{lc}(\phi)$	[4:2:8]	-
$O_{pc}(\phi)$	[4:2:8]	-

V. EXPERIMENTS & RESULTS

In this section the final two stages from the methodology which was introduced in Section IV are discussed. Four types of experiments on accuracy, speed, and fatigue, are conducted in which empirical data is used.

A. Step 6: Data analysis

1) Accuracy: For this experiment, the assumption is made that the operators only differ in their accuracy parameter, while they are identical in all other parameters. Jesus provides an empirical result indicating that 28 % of operators are level of service sensitive (TPR = 0.96, FPR = 0.24), 59 % are highly secure (TPR = 0.95, FPR = 0.08), and 13 % are highly efficient (TPR = 0.45, FPR = 0) [37]. Based on this result, $2/6$ of O_{xr} was set to level of service sensitive, $3/6$ of O_{xr} was set to highly secure, and $1/6$ of O_{xr} was set to highly efficient. The O_{lc} parameters were set identical as the O_{xr}. For the O_{pc}, each operator was set to a different type, thus $1/3$ of each type.

Figure 3 shows the results in terms of hit rate and average queue time for all alternatives. Alternatives range from 332 s queuing time to 376 s, and 0.722 to 0.737 on hit rate.

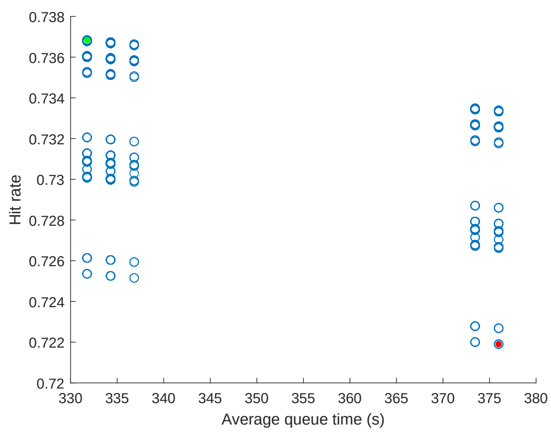


Fig. 3. The result of all scheduling alternatives in terms of hit rate and average queue time. The operators differ only in accuracy. The green alternatives indicates the best, the red alternative indicates the worst alternatives.

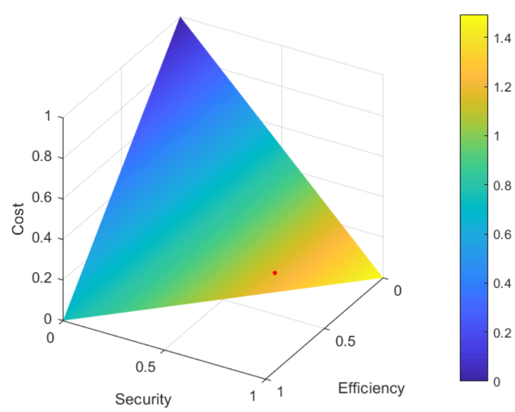


Fig. 4. The difference between the best and worst alternative in terms of aggregated score defined as BW-score, indicated by the colorbar. The operators differ only in accuracy. The axis indicates the weight of the criteria. The red dot indicates the expert weights obtained through discrete choice analysis. BW-score = 1.16 for these weights.

The best alternatives are best in both hit rate and queue time. These alternatives place the highly secure O_{xr} and O_{lc} on the afternoon slot, together with the level of service sensitive O_{pc}. This implies that the best and worst operators in terms of TPR are placed on the morning slot. Putting the efficiency focused operator on the morning slot makes sense, since decreasing average queue time in the morning slot has a larger impact because it affects more passengers. Furthermore, fewer passengers pass per lane in the morning slot compared to the afternoon slot (189 vs 215) , thus the consequence of the low TPR of the efficiency focused operator is lower. It is interesting to find that the level of service sensitive operator is placed on the morning slot. If more passengers pass per lane in the afternoon slot, it would be logical to place the level of service sensitive operator there. However, results that were used to build the surrogate model show that placing a single level of service sensitive operator on the morning slot increases the average hit rate by 0.0039 impacting 755 passengers, whereas on the afternoon slot it increases the average hit rate by 0.0054 impacting 430 passengers. Thus, the total impact on hit rate is bigger in the morning slot. No explanation was found for this.

The worst alternative is worst in both hit rate and queue time. It does exactly the opposite of the best alternative and places one level of service operator and the efficiency focused operator for the O_{xr} and O_{lc} of the afternoon slot, including an efficiency focused O_{pc}.

All the alternatives with an average queue time of 375 s have the highly efficient x-ray operator on the afternoon slot.

Figure 4 shows the difference between the aggregated score of the best and worst alternative. The red dot indicates the weights obtained from experts through discrete choice analysis, seen in Table III. Considering that $\frac{u_1}{w_1} = 203s$ and $\frac{u_2}{w_2} = 0.0157$, a BW-score of 1.16 is quite significant. The BW-score mostly depends on the difference in hit rate, due to its larger weight compared average queue time.

2) Speed: For this experiment, the operators are differentiated on speed. No empirical data was found on how different operators perform in terms of response time. Thus, an assumption is made that one third of operators performs below average, one third performs average, and one third performs above average. The above average operators are assumed to perform 20% better with respect to the average response time, whereas the below average operators are assumed to perform 20% worse. The average response times used are $O_{xr} = 7.5$, $O_{lc} = 64.5$, $O_{pc} = 32.7$.

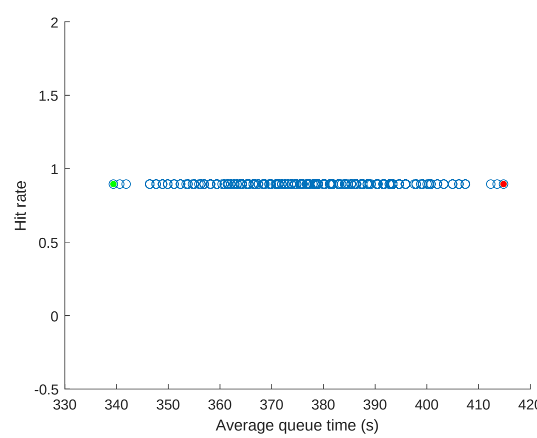


Fig. 5. The result of all scheduling alternatives in terms of hit rate and average queue time. The operators differ only in response time. The green alternatives indicates the best, the red alternative indicates the worst.

Figure 5 shows the different alternatives for the speed experiment. It is seen that the average queue time ranges from $339s$ for the best alternative, till $415s$ for the worst alternative. The worst alternative places the above average performing operators on the afternoon slot, whereas the best alternative places the below average performing operators on the afternoon slot. Since the morning slot has a longer queue and higher average queue time, placing the quicker operators on the morning slot results in a lower average queue time.

The difference between the best and worst alternative is $76s$. Figure 6 shows a BW-score of 0.38 for this difference.

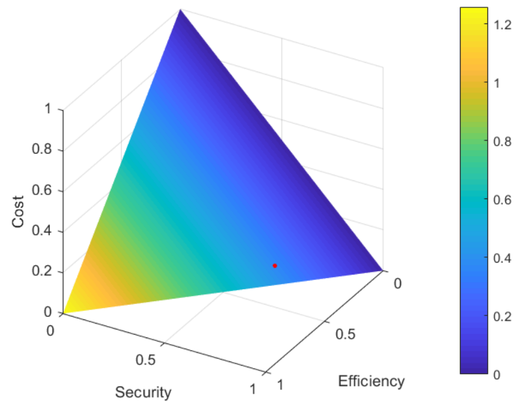


Fig. 6. The difference between the best and worst alternative in terms of aggregated score defined as BW-score, indicated by the colorbar. The operators differ only in response time. The axis indicates the weight of the criteria. The red dot indicates the expert weights obtained through discrete choice analysis. BW-score = 0.38 for these weights.

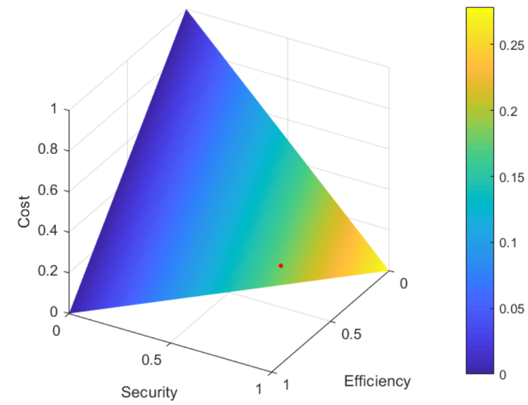


Fig. 8. The difference between the best and worst alternative in terms of aggregated score defined as BW-score, indicated by the colorbar. The operators differ only in fatigue. The axis indicates the weight of the criteria. The red dot indicates the expert weights obtained through discrete choice analysis. BW-score = 0.18 for these weights.

Due to the lower weight of efficiency compared to security, accuracy in operators is more important towards scheduling of operators.

3) Fatigue: For this experiment, the assumption is made that the operators only differ on fatigue. Weighted means of self-reported sleep hours was used as data, which indicates that 7.8% of the sample sleeps ≤ 5 hours, $20.5\% = 6$, $30.8\% = 7$, $41.0\% \geq 8$ [38]. Based on this result, $2/6$ of O_{xr} and O_{lc} set set to 8 hours of sleep, $2/6$ was set to 7 hours of sleep, $1/6$ was set to 6 hours of sleep, and $1/6$ was set to 4 hours of sleep. For the O_{pc} $1/3$ was set to 6, 7, and 8 hours of sleep.

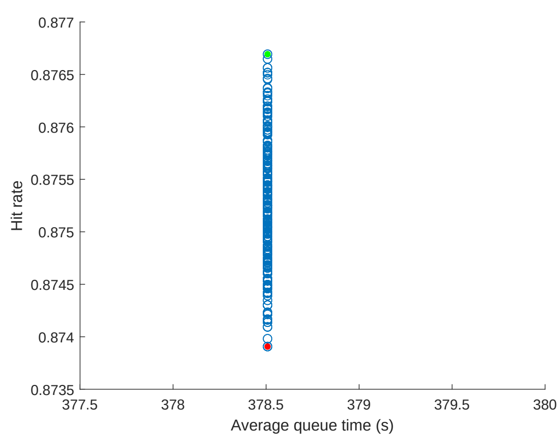


Fig. 7. The result of all scheduling alternatives in terms of hit rate and average queue time. The operators differ in fatigue. The green alternatives indicates the best, the red alternative indicates the worst alternatives.

Figure 7 shows the alternatives of the fatigue experiment.The difference between the best and worst alternative is 0.0028. In terms of BW-score, this is seen in Figure 8. This indicates that fatigue is almost negligible in terms of decision-making, considering the small impact on hit rate and queue time.

4) Combination of factors: For this experiment the operators differ on response time, TPR, and fatigue. A Monte Carlo simulation is performed in which the operator characteristics are specified by probability distributions.

For response time of operators, the assumptions is made that each average response time is multiplied by a random number drawn from $N(1.0, 0.2^2)$. For TPR of operators, 28% is set to $TPR = 0.96$, 59% is set to TPR = 0.96, and 13% is set to $TPR = 0.45$ [37]. For fatigue of operators, 7.8% is set to 4 hours, $20.5\% = 6$, $30.8\% = 7$, and $41.0\% = 8$ [38].

The simulation is run for 1000 configurations, and the results of the difference between the best and worst alternative of each configuration is displayed in Figure 9. The mean BW-score is equal to 0.9788 which is significant considering that $\frac{u_1}{w_1} = 203s$ and $\frac{u_2}{w_2} = 0.0157$. The configurations for which the BW-score is between 0.2 and 0.6 mostly lack an efficiency focused operator (TPR = 0.45), which makes the set of operators indifferent from each other based on TPR. The BW-scores between 0.6 and 1.0 mostly have a single efficiency focused O_{xr} or O_{lc} in the set of operators. Scores above 1.0 mostly have at least two or more efficiency focused operators. This results show that for realistic variations in operator performance, a proper planning tool such as the one proposed in this work, has a significant influence on the performance of the security checkpoint.

B. Step 7: Provide decision-aid

Based on the executed case study, security is significantly more important than efficiency. Therefore, TPR of operators is a more driving parameter compared to their response time. Based on empirical data on TPR of operators, it is seen that significant benefit can be obtained by placing efficiency focused operators on the morning slot peak-hours because they have a large impact on the average queue time caused by the large group of passengers, whereas they are in contact with less passengers compared to the afternoon slot. Even though there

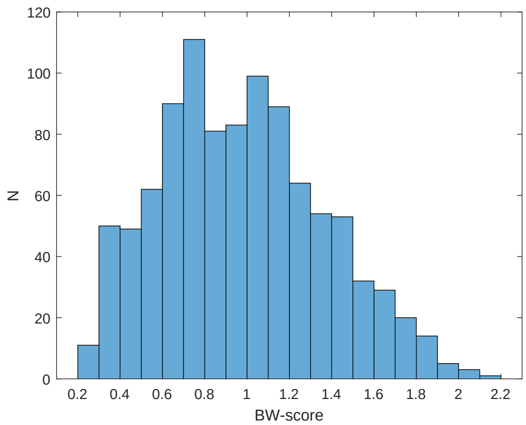


Fig. 9. The difference between the best and worst alternative in terms of aggregated score defined as BW-score, for a total of 1000 different configurations.

is less interaction with passengers in the morning, operators with very high TPR can better be placed on morning slots. The quicker operators can better be placed in the morning slots due to a larger queue. Fatigue can be considered insignificant compared to TPR and response time.

VI. Discussion

The methodology and case study show that integrating multi-criteria decision making methods with agent-based models is promising for decision support for scheduling personnel with increased security checkpoint performance. This section discusses strengths and weaknesses of the study.

A strength of the presented study is that this methodology can be applied to any simulation model for a problem with multiple alternatives and multiple criteria. For example, selecting the best traffic situation considering safety and efficiency in a traffic simulation, or selecting the best evacuation route considering safety and efficiency in a building on fire.

Another strength is the potential practical benefits of the methodology. Considering the case study, if airport managers would measure the performance of their operators, a decision aid tool would allow airport managers to setup their personnel more effectively. The integration provides a structured comparison between alternatives that is intuitive for decision makers.

Another strength is the quantitative support agent-based modelling and simulation brings to MCDM problems. Many MCDM problems depend on expert opinions and qualitative data. Reducing this dependency can lead to more consistent quantitative results which improves decision making.

However, limitations are present in the case study on the agent-based model, surrogate model, the used assumptions, and validation.

A weakness of the case study is the additive translation of the agent-based model to the surrogate model. The used agent-based model was too computationally expensive for the large number of scheduling combinations. Therefore, a good decision support system would require a good surrogate model that is able to replicate the behaviour of the agent-based model with a fraction of the computational cost. With a low computational cost, the methodology could be used for scheduling optimization. However, high-level recommendations, such as in Section V-B, can also be used as input for improving scheduling.

Another weakness in this study is the lack of validation. No accurate data is available on how personnel differs based on response time and TPR. The empirical data used for TPR is not validated data. For response time, assumptions were used. To validate results of the case study, the experiments would have to be tested in reality. Furthermore, for this methodology to be practically used, security operators would have to be tested on their performance.

Other weaknesses exist in the used agent-based model. An advantage of the agent-based model is the use of McCauley's fatigue model which accounts for the effects of sleep on operator performance. However, in both [4], and this study, the effects of fatigue are found to be very low. If the negligible effect of fatigue could be validated, the added value of using the diffusion model is reduced since the diffusion model also calibrates its parameters to real life data. A data-driven statistic approach could improve accuracy of operators in the agent-based model, while reducing complexity and increasing flexibility.

VII. Conclusion and future work

In this paper, the integration of multi-criteria decision making methods with agent-based models was investigated, with the purpose of aiding airport terminal decision makers in configuring personnel to improve security and efficiency. To accomplish this, an existing agent-based model was extended to incorporate differences in performance of individual operators. The agent-based model was used to generate results on average queue time, an efficiency criterion, and true positive rate, a security criterion. Multi-criteria decision making was used to weight average queue time, vulnerability, and average operator cost, and to aggregate a score. The common weighted sum model was used for aggregation of a score, whereas discrete choice analysis was used to determine the weights. A total of 70 security expert observations were collected to accommodate the weightings process.

The results of the case study show that integration of multi-criteria decision making methods with agent-based models allows for a novel practical method of personnel scheduling decision aid for improvement of security and efficiency. The results show that taking performance quality of employees into account during personnel assignment can lead to a better overall security and efficiency. A set of recommendations was identified by which schedules can be improved to improve security and efficiency at airport terminal operations. This includes placing efficient low accuracy operators on morning peak hours because of the larger group of passengers it affects in terms on queue time, whereas it affects less passengers in terms of hit rate due to less passengers per lane. Placing operators with a very high accuracy can also be placed on the morning slots, for reasons which were not identified. The quicker operators can better be placed in the morning slots due to a larger queue. Furthermore, the impact of fatigue was found to be insignificant.

Future work on this topic includes improving on the weaknesses. The translation of the agent-based model to the surrogate model is considered a weakness which can be improved. Furthermore, validation of the performed work is another weakness. Validation through data collection on how operators differ is useful towards this study.

References

[1] S. Greco, J. Figueira, and M. Ehrgott, *Multiple criteria decision analysis: State of the art surveys*. Springer, 2016.

[2] D. Wilson, E. K. Roe, and S. A. So, "Security checkpoint optimizer (sco): An application for simulating the operations of airport security checkpoints," in *Proceedings of the 38th conference on Winter simulation*, Winter Simulation Conference, 2006, pp. 529–535.

[3] S. Janssen, A. Sharpanskykh, and R. Curran, "Agent-based modelling and analysis of security and efficiency in airport terminals," *Transportation research part C: emerging technologies*, vol. 100, pp. 142–160, 2019.

[4] A. Knol, A. Sharpanskykh, and S. Janssen, "Analyzing airport security checkpoint performance using cognitive agent models," *Journal of Air Transport Management*, vol. 75, pp. 39–50, 2019.

[5] K. E. Yoo and Y. C. Choi, "Analytic hierarchy process approach for identifying relative importance of factors to improve passenger security checks at airports," *Journal of Air Transport Management*, vol. 12, no. 3, pp. 135–142, 2006.

[6] A. A. Kirschenbaum, "The cost of airport security: The passenger dilemma," *Journal of Air Transport Management*, vol. 30, pp. 39–45, 2013.

[7] D. H. Harris, "How to really improve airport security," *Ergonomics in Design*, vol. 10, no. 1, pp. 17–22, 2002.

[8] A. Schwaninger, "Increasing efficiency in airport security screening," *WIT Transactions on the Built Environment*, vol. 82, 2005.

[9] J. Skorupski and P. Uchroński, "A fuzzy reasoning system for evaluating the efficiency of cabin baggage screening at airports," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 157–175, 2015.

[10] A. A. Kirschenbaum, M. Mariani, C. Van Gulijk, S. Lubasz, C. Rapaport, and H. Andriessen, "Airport security: An ethnographic study," *Journal of air transport management*, vol. 18, no. 1, pp. 68–73, 2012.

[11] A. Schwaninger, "Airport security human factors: From the weakest to the strongest link in airport security screening.," in *Proceedings of the 4th International Aviation Security Technology Symposium*, 2006.

[12] N. F. Matsatsinis and P. Delias, "Agentallocator: An agent-based multi-criteria decision support system for task allocation," in *International Conference on Industrial Applications of Holonic and Multi-Agent Systems*, Springer, 2003, pp. 225–235.

[13] L. Yu, S. Wang, and K. K. Lai, "An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring," *European journal of operational research*, vol. 195, no. 3, pp. 942–959, 2009.

[14] A. Amato, B. Di Martino, and S. Venticinque, "Agents based multi-criteria decision-aid," *Journal of Ambient Intelligence and Humanized Computing*, vol. 5, no. 5, pp. 747–758, 2014.

[15] S. Janssen, A.-N. Blok, and A. Knol, *Aatom - an agentbased airport terminal operations model*, 2018.

[16] R. Ratcliff and J. N. Rouder, "Modeling response times for two-choice decisions," *Psychological science*, vol. 9, no. 5, pp. 347–356, 1998.

[17] P. McCauley, L. V. Kalachev, D. J. Mollicone, S. Banks, D. F. Dinges, and H. P. Van Dongen, "Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance," *Sleep*, vol. 36, no. 12, pp. 1987–1997, 2013.

[18] J. C. Thiele, W. Kurth, and V. Grimm, "Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using netlogo and r," *Journal of Artificial Societies and Social Simulation*, vol. 17, no. 3, p. 11, 2014.

[19] F. Lamperti, A. Roventini, and A. Sani, "Agent-based model calibration using machine learning surrogates," *Journal of Economic Dynamics and Control*, vol. 90, pp. 366–389, 2018.

[20] T. L. Saaty, "Decision making with the analytic hierarchy process," *International journal of services sciences*, vol. 1, no. 1, pp. 83–98, 2008.

[21] J.-P. Brans, P. Vincke, and B. Mareschal, "How to select and how to rank projects: The promethee method," *European journal of operational research*, vol. 24, no. 2, pp. 228–238, 1986.

[22] B. Roy, "The outranking approach and the foundations of electre methods," in *Readings in multiple criteria decision aid*, Springer, 1990, pp. 155–183.

[23] R. L. Keeney and H. Raiffa, *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press, 1993.

[24] S. Opricovic and G.-H. Tzeng, "Extended vikor method in comparison with outranking methods," *European journal of operational research*, vol. 178, no. 2, pp. 514–529, 2007.

[25] K. P. Yoon and C.-L. Hwang, *Multiple attribute decision making: an introduction*. Sage publications, 1995, vol. 104.

[26] J. Rezaei, "Best-worst multi-criteria decision-making method," *Omega*, vol. 53, pp. 49–57, 2015.

[27] D. Pamučar, Ž. Stević, and S. Sremac, "A new model for determining weight coefficients of criteria in mcdm models: Full consistency method (fucom)," *Symmetry*, vol. 10, no. 9, p. 393, 2018.

[28] A. Guitouni and J.-M. Martel, "Tentative guidelines to help choosing an appropriate mcda method," *Euro-*

pean Journal of Operational Research, vol. 109, no. 2, pp. 501–521, 1998.

[29] M. Velasquez and P. T. Hester, "An analysis of multi-criteria decision making methods," *International Journal of Operations Research*, vol. 10, no. 2, pp. 56–66, 2013.

[30] R. K. Gavade, "Multi-criteria decision making: An overview of different selection problems and methods," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5643–5646, 2014.

[31] R. E. Nisbett and T. D. Wilson, "Telling more than we can know: Verbal reports on mental processes.," *Psychological review*, vol. 84, no. 3, p. 231, 1977.

[32] O. I. Larichev and H. M. Moshkovich, *Verbal decision analysis for unstructured problems*. Springer Science & Business Media, 2013, vol. 17.

[33] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, "Normalization techniques for multi-criteria decision making: Analytical hierarchy process case study," in *doctoral conference on computing, electrical and industrial systems*, Springer, 2016, pp. 261–269.

[34] A. Knol, "Modelling and analysis of dynamic relations and trade-offs between security and efficiency in airport security operations," 2018.

[35] K. Leone and R. R. Liu, "Improving airport security screening checkpoint operations in the us via paced system design," *Journal of Air Transport Management*, vol. 17, no. 2, pp. 62–67, 2011.

[36] S. van Cranenburgh and A. T. Collins, "New software tools for creating stated choice experimental designs efficient for regret minimisation and utility maximisation decision rules," *Journal of Choice Modelling*, vol. 31, pp. 104–123, 2019.

[37] M. Jesus, "Trade-off analysis between security and efficiency of airport operations: A methodological underpinned modelling approach to identify, analyse and evaluate trade-offs based on empirical choice observations," 2018.

[38] P. M. Krueger and E. M. Friedman, "Sleep duration in the united states: A cross-sectional population-based study," *American journal of epidemiology*, vol. 169, no. 9, pp. 1052–1063, 2009.

II

Thesis Book of Appendices

A

LITERATURE REVIEW[1]

This section provides some of the relevant theory from the literature review. Section A.1 discusses the concepts of security and efficiency at the airport terminal. Then, Section A.2 shows the usage possibilities for airport models. Finally, Section A.3 reviews existing MCDM methods and recent trends.

A.1. SECURITY & EFFICIENCY

AIRPORT SECURITY

The importance of improving airport security has been a central aviation theme throughout the years. Security risk management is concerned with assessment and management of security risks to optimize for risk. Risk is synonymous with the potential for loss due to unwanted events, and is the general security indicator.

Security risk assessment is used to identify, estimate, and analyze security risks. Threat, Vulnerability and Consequence (TVC) is a frequently used security risk assessment method to determine risks [1, 2]. Threat, vulnerability, and consequence are the three parameters that describe a security risk. The sequence of TVC is as follows: first a threat is identified, the consequence is assessed, the threat likelihood is assessed, the vulnerability is assessed, then the total risk of the threat is calculated. The probabilities and risks for such attacks are often estimated by security expert knowledge. This is a limitation of the TVC method since it can easily produce wrong probability and risk values [3].

Agent-based modelling is an alternative method, proposed by Janssen, that can improve security risk estimates and reduce dependency of security expert knowledge [4, 5]. A better security risk can be estimated through parameter quantification and using Monte Carlo simulations where intelligent agents are modelled in an airport terminal environment. Expert-based input, such as from the TVC method, is still required to determine inputs to the agent-based model. The advantage of using agent-based models is the ability to take into account spatiotemporal aspects and the complex interaction between passengers and security operators [5].

AIRPORT EFFICIENCY

Besides security, improving efficiency within airport terminals is a major objective. Two different efficiency types exist regarding airports: financial and operational [6, 7]. Financial efficiency depends on revenues and consists of aeronautical revenue (landing, handling, passengers, etc.) and non-aeronautical revenue (duty free sales, retail store, restaurant, etc.). Operational efficiency refers to practical aspects of the airport. For the airport terminal, this mainly includes measures regarding passenger throughput, which describe time and quantity.

TH Airport Consulting carried out a survey among members of ACI Europe Regional Airports Forum to find important operational efficiency indicators. For terminal operations, passenger throughput at security and check-in are two of the most important indicators for operational efficiency [8]. Passenger throughput at border control is also an efficiency indicator, but is less important considered to security and check-in. A key measure to increase operational efficiency is to optimize the security checkpoint processes. Throughput is increased if queuing time is decreased.

[1] The literature review has been previously graded under AE4020.

AIRPORT TERMINAL FACTORS

There are three major factors that affect airport terminal security effectiveness: human factors, facility & equipment, and responsibilities and procedures [9]. Facility & equipment mostly depends on the quality of equipment. Responsibilities & procedures are, for example, the procedures and policies that a contracted screening company follows. Comparing the 3 major factors using the analytic hierarchy process, Yoo found that human factors are the most important factors in the passenger screening process [9]. Yoo identified that problem elements are high labor turnover rate, insufficient training, fatigue due to high work load in peak hours, low wage, and insufficient quantity in human resources. Therefore, Human factors are for example: skill level, fatigue, personality type, security focus, motivation.

Besides Yoo, others have stressed the importance of human factors: it is an often recurring theme is that human factors have a critical role in improving security systems [9–12]. Kirschenbaum states that passenger behaviour is rarely taken into account in airport security design [10]. It is stated that the screening process is technically not supposed to have social meaning. This is not the case as there is both social meaning and social interaction between employees and passengers [13]. An aspect of social interaction is negotiation between employee and passenger. It is stated that four general behaviours exist at the security check point: (I) the passenger passes through the security checkpoint without issues, (II) the passenger accepts to remove carried forbidden items, (III) the passenger will negotiate with the employee to try and bring along the forbidden item, (IV) the passenger refuses or argues to comply [10]. It is stated that the cost of providing security is increased substantially by passengers that negotiate. Furthermore, as stated before, security employees often bend rules in 20 to 40 % of decisions that are made, which contributes to the fact that employees do not always follow procedures and protocols.

Harris attempts a scientific assessment to identify how airport security could be improved [11]. Harris mentions that improving one specific component leads to exploitation of weaknesses in other components. It is argued that effective security personnel performance is affected by the design of operator task, the task environment, selection of operations, and training of operators. Furthermore, the security operators should be placed in positions which matches their capabilities and aptitudes. Increase in pay could potentially prevent personnel to quit, which could indirectly lead to increased performance. However, pay is not a direct reflection of how performance could be increased. Harris concludes by mentioning that airport security should change by becoming smarter and more flexible, and not necessarily more extensive. However, it is not explained or suggested how this can be achieved.

Schwaninger states that good equipment is of limited value if the operating personnel does not perform their task accurately and efficiently [12]. The conducted study tested the performance of x-ray operators by assessing their detection performance in various situations. Different viewpoints, super-positions, and bag complexity situations were addressed with guns, knives, and IEDs as forbidden items. An individual adaptive training proved to be an useful tool which has increased the detection performance speed of the operators by almost 50 %, after six months of training.

There are different possibilities to investigate human factors. For example, Schwaninger divides screener performance among four determinants: aptitude abilities, visual knowledge, motivation & attention, and correct action [14]. However, it is possible to also look at other criteria such as education, previous experience, and health [15]. Agent-based modelling is a suitable paradigm to investigate human factors since the behaviour of the agents can be altered to represent security operators with different skill sets.

A.2. AIRPORT TERMINAL MODELS

Airport terminal model usage can be categorized in four types: capacity planning, operational planning and design, security planning and design, and airport performance review [16]. An other more general way of categorization is with 3 classes: explorative, predictive, and operative [17].

Capacity planning models are used to test infrastructure within the airport environment regarding their capacity. Capacity planning falls within the predictive category because it answers a "what if?" question [16, 17]. Simple Landside Aggregate Model (SLAM) is one of such models which can test alternative configurations of processing and holding facilities [18]. Other simulation models that focus on airport capacity and infrastructure are pedestrian models like Pedestrian Dynamics by INCONTROL [2] and NOMAD by TU Delft [19].

Operational planning and design models are concerned with day-to-day planning such as resource allocation, load testing, and efficiency testing. For example, number of service counters that have to open to handle the expected incoming passengers. The same inputs and outputs are often used as with the capacity planning models. Operational planning and design models often use agent-based modelling. Operational planning falls within the operative category because it is concerned with optimized solutions to operational problems. An example of a model capable of operational planning is the security checkpoint optimizer by Wilson, which allows for security effectiveness tests, throughput tests, and resource allocation [20].

Security planning and design models are concerned with modelling security in airport terminals. Security risk, a function of threat, vulnerability, and consequence, is likely to be an output which is optimized for in these models. Security models are used as aid to test new technologies or security policies. A difficulty of creating security models is obtaining large amounts of useful data since threat events are infrequent. Simple probability or Bayesian models can be used for security planning and design. More detailed models require agent-based modelling to capture the complexity of the problem. Security planning can fall within the predictive category because alternative planning configurations and policies can be tested. However, security planning can also address optimized solutions to operational problems. An example of a security planning model is Perboli's AirSim [21], which is used to simulate the system performance of new policies. An other example of a security planning model is the risk model by Chawdry in which the risk of a new security policy is assessed [22]. The security checkpoint optimizer by Wilson can also be used for security planning, as it allows for evaluation of policy changes.

Airport performance review models are used to estimate passenger levels of satisfaction. A selection of performance metrics is made which have influence on passenger satisfaction. Waiting and service time have considerable effect on passenger satisfaction. Performance review models are subject to subjectivity as individuals differently value performance metrics. Airport performance review models fall within explorative category because explorative models correspond to fuzzy qualitative system reviews. An example of an airport performance review model is the model by Tsaur which is used for evaluation of airline service quality [23]. A summary of the four types of model usages is seen in Table A.1.

Table A.1: Airport terminal model usage summary [16]

Model Type	Model Class	Usage
Capacity planning	Predictive	Future design changes
Operational planning & design	Operative	Day-to-day resource allocation
Security planning & design	Predictive/Operative	Test technology/security policy
Airport performance review	Explorative	Passenger satisfaction

Even though separate categories can be defined for airport terminal models, this does not imply that some models can't be used for multiple purposes. For example, an agent-based model running a simulation where passenger agents interact with security agents could be used for multiple purposes, such as Wilson's security

[2] https://www.incontrolsim.com/, retrieved on 12/11/2018

checkpoint optimizer which can be used for capacity, operational, and security planning problems [20].

AGENT-BASED MODELS

Regarding level of detail, there are different types of models that can be used to model security and efficiency in the airport terminal. A broad model categorization on level of detail is macroscopic, mesoscopic, and microscopic [16]. Macroscopic models are the simplest type of models as they don't incorporate interaction between elements. These models attempt to describe the average values of the indicators of the whole system. Examples of macrosopic airport terminal models are queue models. Macroscopic models are unable to handle the complex airport terminal environment and therefore are not suitable for accurate modelling. Mesoscopic are more detailed than macroscopic models. They however lack detailed passenger interaction.

As computational power increased throughout history, level of detail of models increased. Agent-based models is a class of models that simulates individual or collective interactions of autonomous agents to analyze the system. Capturing individual interactions requires a high level of detail and large amounts of data, and therefore falls under the microscopic model type. If security and efficiency indicators of airport terminal operations are to be analyzed, agent-based models are most suitable as they can simulate the complex interaction between passenger and security agents, simulate passenger behaviour and movement, and the airport terminal space.

An agent-based tool which analyzes multiple criteria including security and efficiency is presented by Wilson [20]. The software is designed to perform "what if" analyses to balance resources and therefore falls in the predictive usage type. The goal of the software is to evaluate potential policy impacts. For example, policy changes such as increasing random carry-on item searches or a change in prohibited item list will change the parameters in the agent-based model which results in different performance indicators. Furthermore, Wu presents a review of models and describes several agent-based models [16]. As seen in Wu's review, most of the described agent-based models are used for operational planning.

AGENT-BASED OPTIMIZATION APPLICATIONS

A conducted literature review on agent-based model application for optimization problems found that most agent-based optimization applications tackle scheduling problems [24]. Real world application in the aerospace industry is minimal.

Personnel assignment regarding multiple criteria has been explored in various situations. One paper describes the use of Analytic Hierarchy Process (AHP) to generate competence profiles for military personnel [15]. The goal is to match profiles with positions, where position preferences of the personnel is also taken into account. An other study uses fuzzy AHP and Analytic Network Process (ANP) to match candidates and positions for multiple objectives [25].

Both of these studies use a qualitative comparison of human factors to find the best suiting personnel. A better way to approach this is to quantify the impact of human factors, and use simulations models to obtain a quantitative output such that a more accurate score can be aggregated. This is where agent-based modelling can be used to observe performance indicators. Two papers describe a multi-agent based approach for personnel scheduling in assembly centers [26, 27]. Their goal is to minimize two objectives: operational cost and personnel dissatisfaction. Their agent-based model utilizes distributed problem solving in which agents negotiate to form coalitions and improve their individual schemes and the global solution.

A possible option is to vary parameters of personnel in the agent-based model to observe the resulting security and efficiency indicators. Some of these factors are fixed by the human resource competencies (e.g. personality, aptitude), whereas some factors can be changed by the human resource itself (e.g. experience, training). A single set of human factors can be the input to the agent-based model. Observations of the agent-based model then results in obtaining security and efficiency indicators. MCDM methods can then be used to weigh the different indicators and obtain a score such that the decision maker can make a selection between possible alternatives.

AATOM

As mentioned before, this study revolves around how agent-based models can be used to aid airport decision makers in making decisions that concern both security and efficiency of airport terminal operations. This implies that agent-based modelling itself is not part of the study or thesis. Instead, the focus is on decision aid for airport operators through use of agent-based models. This study continues research within TU Delft its aerospace faculty and was preceded by work from Stef Janssen concerning security and efficiency in airport terminals [4, 28]. Other MSc students have made contributions to topics of security risk assessment [29], resilience [30], quantifying vulnerabilities [31], and trade-off relations between security and efficiency [32, 33]. Through all preceding work, agent-based security risk management models have established a framework to explore uncharted territory. The preceding work by TU Delft has been done in collaboration with Rotterdam The Hague airport.

The model that has been developed by Stef Janssen is an Agent-based Airport Terminal Operations Model (AATOM) [34]. The model contains the main handling processes of outbound passengers: check-in, security, and border control. The baseline model of AATOM can be used for different kind of studies related to security, efficiency, resiliency, and safety. The baseline model allows to specify passengers parameters, operators parameters, flights, and sensor parameters. A full description of input parameters, and the assumptions made in AATOM can be found in the paper [34]. The AATOM baseline model serves as a good starting point to quantitatively extract security and efficiency indicators.

A.3. MULTI-CRITERIA DECISION MAKING

The general framework for a MCDM method can be seen in Figure A.1. Input is the obtained data that is used for comparison. For example, an agent-based model can generate data like queue length and queue time to be used as input. The modelling process consists of applying the MCDM technique to determine weights and scores. Aggregation follows the modelling, after which a recommendation can be made based on different scores.

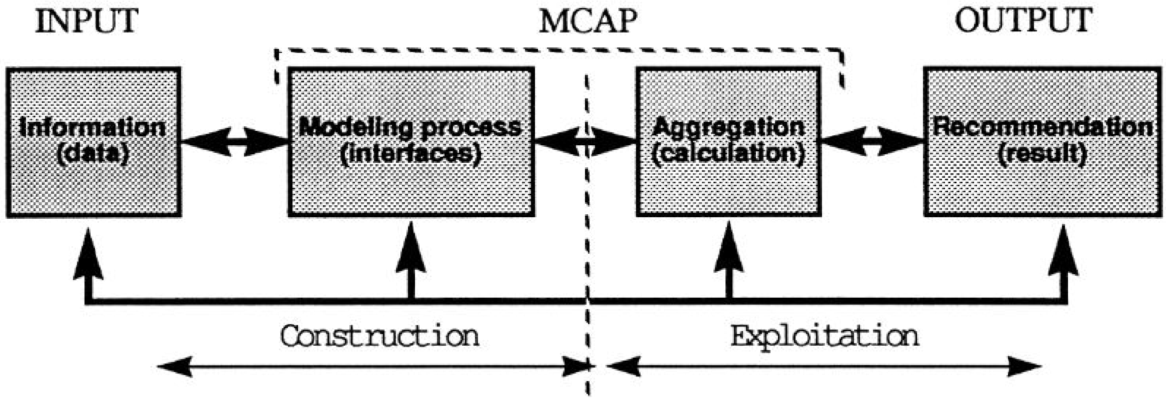


Figure A.1: General framework for a MCDM method [35].

In general, the problem statement is formulated as seen in Equation A.1. Here a_j represents an alternative, c_i represents a criterion, and q_{ij} represents the score of the alternative i respect to criterion j.

$$A = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_i \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_j \end{array} \\ \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1j} \\ q_{21} & q_{22} & \cdots & q_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ q_{i1} & q_{i2} & \cdots & q_{ij} \end{pmatrix} \end{array} \tag{A.1}$$

The overall score of an alternative, V_i, is the summation of all scores with respect to its criteria. Since the criteria have varying importance, they are assigned weights, w_j, where $\sum w_j = 1$ and $w_j \geq 0$. Thus an overall score can be represented as Equation C.1.

$$V_i = \sum_{j=1}^{n} w_j q_{ij} \tag{A.2}$$

The difference between different MCDM methods is the way in which the weights are distributed.

POPULAR MCDM METHODS

This subsection explores the most popular used MCDM methods. It will list and explain the steps that are taken with each method. Furthermore, advantages and disadvantages are mentioned. The technical details are not stated due to irrelevance at this stage. Having basic knowledge about the main MCDM methods is important to understand the core of MCDM, and understand differences between the methods.

MAUT

MAUT is one of the methods to make a choice concerning a set of alternatives. It involves trade-offs between multiple attributes within the system. For example, average queue time and vulnerability can be two airport terminal attributes that provide a certain utility towards the decision maker. If the decision maker can reduce the passenger queue time level by increasing vulnerability level, the decision maker needs to know what combination of attribute levels provides the largest utility. Utility is a numerical function that quantifies the preference of the decision maker. By defining utility, all potential alternatives can be ranked from most desirable to least desirable. Preferences can be compared with each other in a binary way where two attributes are indifferent, or one attribute has weak or strong preference. However, within utility theory, preference modelling is done numerically by defining utility functions. The total utility can be determined through direct or decomposed assessment. A direct assessment directly estimates the combined utility of all attributes, whereas a decomposed assessment estimates the utility of each single attribute and then combines it.

Construction of the utility function for multiple attributes becomes difficult because complexity increasingly increases with each attribute. A large amount of precise input data is required to accurately model the decision maker preferences. This is disadvantage of the MAUT method, since accurate input data for the problem is often not available [36]. An advantage of MAUT is that uncertainty is accounted for. Therefore, the method is often used with problems that have high uncertainty and large amount of data.

SMART is the simplest MAUT based MCDM method. Compared to most MAUT methods, SMART doesn't require judgments of preference or indifference between the potential actions [37]. An advantage of SMART is its ease of use. The method does oversimplify the problem which can lead to the top alternatives being indistinguishable [38].

AHP

AHP is an other popular MCDM method. The main feature of AHP is pairwise comparison of criteria weights and potential actions. AHP consists of three steps:

1. Computing criteria weights
2. Computing option scores
3. Ranking options

In the first step, the criteria weights are computed using the pairwise comparisons, and assigning a value. For example, if two criteria are equally important the score is a 1, if one criteria is slightly preferred then the score is a 3, if it is preferred the score is a 5, if it is truly preferred the score is a 7, and absolute preference a score of 9. These values build up a matrix A. Then, A is normalized and the entries on each row are averaged, which results in obtaining the criteria weight vector w. In the second step, a matrix B is constructed where different options (potential actions) are compared with respect to the criteria. Again pairwise comparisons are used to build a matrix, B. Then, B is normalized and the entries on each row are averaged, which results in obtaining the score matrix S. Finally, a vector of global scores is obtained by $v = S \cdot w$. Now the scores can be ranked according to their score. Furthermore, AHP allows a consistency test to be used to estimate if errors have occurred while modelling the pairwise relations.

AHP is the most frequent applied MCDM method found in literature [39]. Advantages of AHP are its ease of use, scalability, and adjustability [36]. AHP requires less data input than MAUT, and can be applied to larger problems. However, for large problems the number of pairwise comparisons can become large $(n(n-1)/2)$. A disadvantage of AHP can be interdependence between criteria and alternatives. Rank reversal might also be possible if a similar alternative is added to the set of alternatives that is used for evaluation.

Analytic Network Process (ANP) is an other method which was developed from AHP. ANP is a more general form of AHP which structures the problem as a network and allows for interdependencies, outerdependencies, and feedbacks between decision elements [40]. Compared to AHP, ANP can be a better tool to gain more information about the relation between elements. Furthermore, some problems can only be described by ANP. ANP is much less popular then AHP because it is more complex than the simple AHP. Due to the dependencies and feedbacks, verification is not possible in ANP.

TOPSIS

The concept of TOPSIS is that the best potential action should have the shortest distance to the positive ideal solution, and the farthest distance from the negative ideal solution. TOPSIS consists of the following 7 steps [41]:

1. Construct decision matrix and compute criteria weight
2. Normalize decision matrix
3. Calculate weighted normalized decision matrix
4. Determine positive ideal and negative ideal solution
5. Calculate separation measures of each alternative
6. Calculate relative closeness towards positive ideal solution

7. Ranking alternatives

The decision matrix contains the potential actions, and the criteria weights are directly divided among the multiple attributes such that all criteria weights add up to 1. Normalization of the decisions allows for comparison. The ideal solution is found which maximizes the benefit criteria and minimizes the cost criteria, whereas the negative ideal minimizes benefit and maximizes cost. From here, the distance is measured from these ideals to determine a relative closeness value. The relative closeness value is used to rank the alternatives.

An advantage of the TOPSIS method is its ease of use and the ability to be programmed [36]. A disadvantage is that the use of Euclidean distance disregards attribute correlation.

VIKOR

VIKOR is another method in which the aggregating function represents closeness to the ideal solution, just as TOPSIS. The approach to the problem is similar to TOPSIS, with some minor differences. Opricovic states the main differences between VIKOR and TOPSIS [42]:

- VIKOR uses linear normalization whereas TOPSIS uses vector normalization.
- VIKOR its solutions provides maximum 'group utility' for the 'majority' while minimizing 'regret' for individual criteria.
- VIKOR its highest ranked alternative is closest to the ideal solution, whereas TOPSIS its highest ranked alternative is performs best in terms of the ranking index, which does not necessarily mean that it is closest to the ideal solution.

PROMETHEE

PROMETHEE is a family of MCDM methods from the outranking category. PROMETHEE I is a partial ranking method, PROMETHEE II a complete ranking method, PROMETHEE III considers ranking based on interval, PROMETHEE IV considers a continous set of alternatives, PROMETHEE V considers problems with segmentation constraints, and PROMETHEE VI considers human brain representation [43]. To explain the method, the procedure for PROMETHEE II will be explained as the steps in this method are fundamental in the procedure of the other PROMETHEE methods. Furthermore, PROMETHEE II is mostly referred to by the scientific community [43]. The method consists of the following 5 steps:

1. Compute deviations
2. Apply preference function
3. Calculate global preference index
4. Calculate outranking flows
5. Rank alternatives

The first step consists of computing deviations which implies that pairwise comparisons are made between the criteria to assign values of importance relative to each other. The preference function is then applied to determine what type of preference there is between the criteria. For example, weak preference, strong preference, or absolute preference. Six different preference functions were defined by the PROMETHEE pioneers [44]. Then, the global preference index, $\prod(a,b)$, is calculated by summing all criteria with their respective weights. It represents the intensity of the preference. The global preference indices are summed to obtain the outranking flows. The net flow is equal to the leaving flow minus the entering flow. A higher net flow implies a better alternative.

A PROMETHEE method advantage is its ease of use. A disadvantage of this method is that it does not provide a structured way to determine values of the weights [36].

ELECTRE

ELECTRE is a family of MCDM methods from the outranking category [45]. ELECTRE I is a choice problematic method. ELECTRE II, III, and IV are ranking problematic methods. ELECTRE I is not able to rank the methods and does not account for uncertainty. Ranking ability was incorporated in ELECTRE II, and uncertainty was incorporated in ELECTRE III. ELECTRE IV was developed for situations where importance between criteria can not be quantified. The first general steps for ELECTRE are [46, 47]:

1. Construct decision matrix
2. Normalize decision matrix
3. Calculate weighted normalized decision matrix
4. Calculate concordance and discordance indices

The first three steps are the same as in the TOPSIS method. For each criterion preference threshold, indifference threshold, and veto thresholds, and importance rating have to be determined. This is a process where an attempt is made to quantify the criteria. More information about thresholds can be found in [45]. Step 4 determines the concordance and discordance indices. The indices are calculated by the threshold and rating performance measures. Concordance and discordance are relative performance measures between alternatives. After index construction, the outranking relations are constructed. This procedure is different per ELECTRE method. [45] provides further information and sources on ELECTRE methods.

RANK REVERSAL

A problem many decision making methods can face is rank reversal. This means that the ranking of the alternatives changes if a different method is used, or by addition of a new alternative to the existing set of alternatives. For example, consider a MCDM method is applied and the output is 3 alternatives ranked where $A > B > C$. Now if an alternative D is added to the set of alternatives, and it is known that D is a worse alternative than C, then it should be that $A > B > C > D$. However, it is possible that now, for example, $B > A > C > D$ which means that rank reversal occurred. This is a problem within MCDM because it is impossible to determine if the 'best' alternative has been found. Furthermore, different methods can give contradicting or different results. Therefore, it is not possible to determine the best method without knowing the best method in advance [48]. This is known as the decision-making paradox.

MCDM SELECTION GUIDELINES

There is a wide variety of MCDM methods where each approach is different, and not a single approach exists that is able to be compatible with each situation. Guidelines towards choosing a single MCDM method exist, but it is not possible to agree to a universal general method for selecting MCDM methods. When choosing a MCDM technique to solve a particular problem, problem solvers often select a technique they are familiar with, or a technique which has software to support it [49].

Kornyshova analyses and compares nine different selection approaches for selecting a MCDM method [49]. It concludes that each selection approach has its advantages and disadvantages that may change depending on the context. Furthermore, Kornyshova suggests that a better approach of selecting a MCDM method is possible and gives 12 new guidelines which includes: (I) Take into account problem situation, (II) Allow a typology of problem characteristics, (III) Consider MCDM techniques specifics, (IV) Take into account data diversity (types, scales etc.), (V) Consider all main groups of MCDM techniques and be able to deal with a new one, (VI) Present a more precise estimation for parameters as alternatives number and easiness of use, (VII) Allow selecting of MCDM technique, as well as its better understanding and adaptation to a concrete case, (VIII) Take into account interaction between goals, (IX) Be structured, (X) Be universal as regards to application domain, (XI) Permit a capitalization of selection results, (XII) Suggest a tool facilitating MCDM techniques selection.

Reflecting on these guidelines, it is seen that a "best" selection of a MCDM method is a difficult and unclear problem. Furthermore, selecting a MCDM method is a paradox, since it requires a MCDM technique to compare different criteria of the MCDM techniques.

Guitouni his conceptual framework has often been used as general guidelines for choosing between MCDM methods [35]. The applicability of 29 different methods is compared based on the problem that is approached. After their analysis, a set of 7 guidelines is suggested to aid MCDM method choice. The guidelines consist of (I) Determine stakeholders, (II) Consider decision maker cognition, (III) Determine decision maker problematic, (IV) Choose aggregation procedure that can handle the input information, (V) Consider the compensation degree, (VI) The fundamental hypothesis of the method is to be verified, (VII) Take into account the decision support system coming with the method.

Velasquez compares the advantages and disadvantages of 11 popular MCDM methods [36]. Furthermore, an analysis is performed on the areas of application for each method. The advantages, disadvantages, and application area, can be used as guidelines for selecting an approriate MCDM method.

MCDM PROGRESS & NOVELTIES

It is evident that all MCDM have some common ground. However, each method has its own way of quantifying the criteria of the problem. The reason there is not a single solid method is because, as discussed in item A.3, it is not possible to determine what the best method or best alternative is. Some methods are more applicable in a certain application area. For example, MAUT has a lot of application in economic, financial, water management, energy management, and agricultural problems, because these problems face uncertainty and have enough available data [36]. This implies that some methods can provide better solutions if the method is compatible with the problem.

As mentioned before, all of these 'main classical' MCDM methods have been extended, changed, and adapted to improve the quality of the decision making process. As discussed before, one of the issues with MCDM is that it is not possible to be certain that the best alternative is chosen. This is because there are numerous uncertainty factors within the problem. If uncertainty can be reduced, the solution is more accurate, and the quality of decision making can increase. The purpose of this section is to discuss how MCDM methods have been improved, and present some recent novel concepts that have emerged. A major researcher in the decision information analysis field was Lotfi A. Zadeh which has introduced the concept of fuzzy set theory, and more recently concepts such as Z-numbers and stratification.

FUZZY SETS

Combining fuzzy set theory with MCDM methods has often been done to address uncertainties within the problem [50]. It is not a novelty since the method has been around for a while. However, almost all MCDM methods have been combined with fuzzy theory in some applications. Fuzzy set theory is a modelling tool to handle uncertain, vague, and imprecise information input. In decision making problems, it is often not the case that the goal and criteria can be well defined through precise information.

A fuzzy set is defined by membership functions. Membership functions indicate the 'degree of truth' of a value to that set. The concept can be explained through Figure A.2.

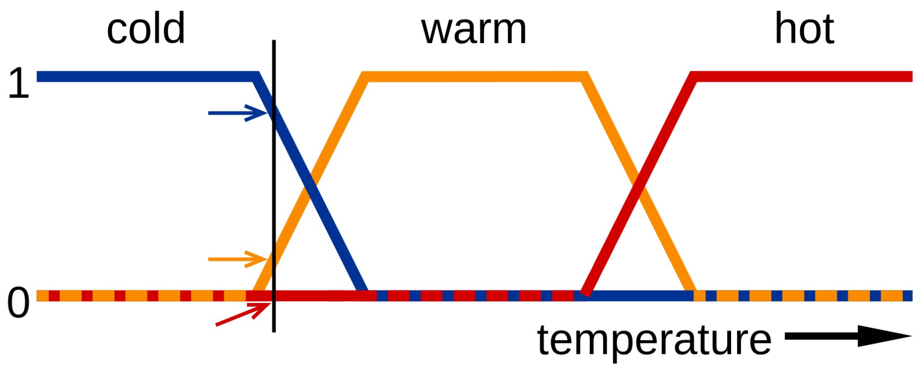


Figure A.2: Fuzzy set of temperatures [3]

The x-axis shows temperature values whereas the y-axis displays the membership value. The vertical line in the figure can, for example, be 8°C. The membership value of the linguistic variable cold would then be 0.8 according to the defined membership function, whereas the value for warm is 0.2. To establish a membership function, a survey can be used to determine what the ranges of values are for the linguistic variables. If a certain variable is described through fuzzy set theory, the final step is often to convert the fuzzy numbers back to a regular 'crisp' number for quantification.

Z-NUMBERS

Z-numbers is another concept dealing with uncertainty of the problem. In comparison with fuzzy sets, Z-numbers are more recent, and allow for better human brain representation. The representation includes restraint (constraint) and reliability (probability) which make op the two components of the Z-number Z(a,b) [51]. Often, Z-numbers are expressed in a natural language. For example, Z(30 min, very sure) can indicate the restraint of an arrival, whereas very sure indicates the reliability.

[3] https://en.wikipedia.org/wiki/Fuzzy_logic, retrieved on 04/12/2018

STRATIFICATION

Stratification is a system of competition where object of competition are strata of data [52]. Stratification means layer formation. A system has a collection of state variables which represent the state of the system. A change in the state variables can result in a state transition. A state-transition function is defined as Equation A.3. Here, u_t indicates input and s_t the current state. Each input results in an output as shown in Equation A.4.

$$s_{t+1} = f(s_t, u_t) \tag{A.3}$$

$$v_t = g(s_t, u_t) \tag{A.4}$$

The concept of stratification has been brought together with the MCDM domain. It addresses the possible fluctuation of criteria weights due to an uncertain dynamic environment. Asadabadi's example in stratified MCDM paper is concerned with possible incidents that could happen in the future that can influence the weighting [53]. It uses the general MCDM problem formulation as seen in Section A.3. However, now each state, k, has its own set of criteria weightings denoted: $Wt_k : \{wt_k, \ldots, wt_{km}\}$. The weight of criterion j in state f is denoted wt_{fj}. A transition matrix is defined, as seen in Equation A.5.

$$P = \begin{array}{c} \\ w_1 \\ w_2 \\ \vdots \\ w_h \end{array} \begin{array}{c} \begin{array}{cccc} w_1 & w_2 & \cdots & w_h \end{array} \\ \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1h} \\ p_{21} & p_{22} & \cdots & p_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ p_{h1} & p_{h2} & \cdots & p_{hh} \end{pmatrix} \end{array} \tag{A.5}$$

The matrix includes the transition probabilities between states. The score of an alternative can then be calculated as seen in Equation A.6.

$$V_{ai} = \sum_{t=1}^{m} q_{it} \sum_{j=1}^{h} wt_{jt} p_{kj} \tag{A.6}$$

The concept of stratification has potential to be applied in airport terminal decision making problems. Security and efficiency weights are subject to change due to the dynamic terminal environment. For example, efficiency weights can become more important to the decision maker if a queue is long, whereas security weights can be more important with a short queue. Probabilities between states could be determined through observation of agent-based simulations. A disadvantage of the method is the increased number of calculations that have to be made. However, uncertainty is reduced on the score of the alternatives.

BWM

One of the newer MCDM methods gaining popularity is BWM [54]. The method uses identification of the best and worst criteria such that less pairwise comparisons can be performed. A pairwise comparison between the best and worst criteria, best and other criteria, worst and other criteria, is required to determine all weights. The steps for BWM are:

1. Determine decision criteria set
2. Determine best and worst criteria
3. Determine preference of best criteria over other
4. Determine preference of other criteria over worst
5. Determine optimal weights

First the criteria are defined which should be used in the decision making process. Next, the best and worst criteria are determined. After this, all criteria can be compared to the best and worst criterion to assess performance over each other. Finally, the optimal weights are found by minimizing the maximum absolute differences.

The advantage of BWM is that it requires $2n-3$ comparisons, which is fewer than $n(n-1)/2$ by AHP. Furthermore, BWM weight derivation has a high reliability and could be incorporated in other MCDM methods [54]. Rezaei proves high reliability by defining some evaluation measures for the results such as consistency ratio, minimum violation, total violation, and conformity. Consistency ratio measures the reliability of the output. Minimum violation measures violations of the ordinal preferences. Total deviation measures actual Euclidean distance between ratios of weights and their comparisons. Conformity measures the performance of the MCDM method by comparing it to intuitive scores. On all the defined evaluation measures, BWM outperforms AHP.

FUCOM

An even more recently defined method is FUCOM [55]. The approach of FUCUM consists of the following steps:

1. Define and rank criteria
2. Comparison of criteria
3. Calculation of weights

In the first step the criteria are simply ranked from most important to least important. Next, the comparison of criteria is carried out where the comparative priority between criteria is determined. The last step is the calculation of weights. The calculation of weight in this method has to adhere to two conditions. The first condition is that the ratio of weight coefficients is equal to the comparative priority. The second condition is that the final values of weights should satisfy the mathematical transitivity conditions. This ensures that a consistent set of weights is obtained.

The method is compared to both AHP and BWM and performs better based on consistency and reliability of the weights. An other advantage of FUCOM is that it only requires n-1 which is even less comparisons than BWM. A disadvantage could be that the implementation is slightly more complex than, for example, BWM. Since the FUCOM method is so new, it has not yet been applied by the scientific community. Time will tell if its popularity will increase.

REFLECTION

A brief look into some of the most popular MCDM methods has revealed that changes exist in the way the methods handle preference information. With the extensions and adaptions included, there are at least 100 different existing MCDM methods. As mentioned in item A.3, the selection of a MCDM method is a decision problem itself. item A.3 also describes some guidelines developed by the analyses of different researchers, but these guidelines are different from each other.

Furthermore, there are many more papers in which researchers have compared and analyzed MCDM methods under different conditions. Saaty, one of the most prominent MCDM researchers and architect of AHP/ANP, reviews scientific literature on the comparative analysis between MCDM methods [56]. Many different conclusions are drawn by applying several MCDM methods to different problem domains. Some conclusions show that for most problems there is little difference in results between different applied MCDM methods, however in a few problems the set of best choices can vary greatly [57, 58].

Saaty concludes that a major problem of different MCDM methods is that they can give different results when applied to the same problem. Another problem is that there is not a 'super method' which is applicable to each decision problem. Furthermore, Saaty suggests that it is difficult to classify, evaluate, and compare MCDM methods as they handle preference information differently. Therefore, 16 criteria are defined to be used for comparing different MCDM methods. It is suggested that AHP is one of the most appropriate MCDM methods available. Furthermore, the very few papers that deal with MCDM methods in airport security and efficiency operation context, AHP or an AHP adaptation is applied [9, 59, 60].

While AHP is the MCDM that is most in use [39, 61], BWM is a new method that performs better. Rezaei, uses consistency ratio, minimum violation, total violation, and conformity as evaluation measures to compare AHP with BWM [54]. This comparison is not performed between the methods, but rather the results the methods produce. However, since this method quite recent, it might be useful to compare it with other MCDM methods to improve BWM its validation. It is secure to use several MCDM methods to be certain that the alternative ranking is of high quality, also mentioned by Salminen [58]. FUCOM is another method similar to AHP and BWM, but seems to provide more reliable weights with less comparisons. The method has not been used by the scientific community yet, and therefore requires more implementation cases for validation.

Furthermore, if required during the decision making process, several adaptions can be performed to reduce uncertainty within the problem. Applying fuzzy set theory is a common example, and Z-numbers are a more recent concept. A Z-number adaptation of the BWM has already been defined to address uncertain information [62]. Stratification can also be applied to account for variation in weights. It is to be noted that fuzzy set theory and Z-numbers will probably not be applied on the crisp output indicators, since the measurement of the indicator is not uncertain.

It can be concluded that a MCDM method is required to evaluate a set of alternatives (potential actions). For the established research question, it is known that these alternatives will explore human factors of personnel in an operational context. Simply, a method like AHP could be applied to establish preference relations between criteria, to determine weights and rank the alternatives. However, it will never be possible to be 100 % certain that ranked alternatives are ranked in the correct order. For the decision maker, a minimization of uncertainty is desired. Therefore, AHP, BWM, and possibly FUCOM, should all be applied while their performance and consistency is measured.

B

Method: Agent-based Model

B.1. Security Screening process

The main airport terminal component regarding security and efficiency is the security checkpoint. The standard measures of a security checkpoint require passenger screening which is performed by employees operating equipment. The basic equipment consists of X-ray scanners and metal detectors. Alternative screening devices, machines, or measures exist, but will not be part of the case study.

There are multiple security officers with different roles present at the security checkpoint. The first security officer that a passenger can encounter is a directions officer. The directions officer does not have a role of detecting forbidden items. The role includes providing support to passengers by giving them directions, instructions, and answers, such that passengers who are unfamiliar with the procedures do not obstruct the passenger flow. Therefore, having a directions officer affects parameters such as the mean time for a passenger to drop its belongings, but possibly also x-ray forbidden item hit rates. Due to the simplicity of the task, human factors of the directions officer are not going to affect the performance. It is unlikely that the performance of a direction officer can be better or worse compared to an other direction officer, to the extent in which it would affect the security and efficiency parameters. Therefore, the directions officer is of little importance regarding the research objective.

When passengers place their luggage on the conveyor system, they will have to move through a Walk Through Metal Detector (WTMD). If the detector gives a positive signal, the physical check officer will have to pat down the passenger. If an item is found during the pat down, the physical check officer reasons if the item has to be confiscated and if the passenger has to move through the WTMD again. Contrary to the directions officer, human factors of the physical check officer do affect security and efficiency parameters. For example, if a physical check officer is quicker in performing the pat down than an other physical check officer, this has a direct impact on the passenger flow, thus an impact on efficiency. Furthermore, there can be differences between officers in the accuracy of the pat down, which will impact the security performance.

When the luggage moves through the conveyor system the luggage will have to be checked by the X-ray officer. The X-ray screening is the main component of the screening process. When inspecting the X-ray image, X-ray officers require abilities such as mental rotation, figure-ground segregation, and pattern recognition to judge the luggage contents. If there are no forbidden items in the luggage, the luggage is cleared. If there are forbidden items in the luggage, the luggage is sent to the bag checker officer for further inspection. If there are doubts about the content of the luggage, the luggage is also sent to the bag checker officer for further inspection. Human factors of the X-ray officer impact the security and efficiency parameters. For example, the mean time an X-ray officer takes to check a bag is approximately 6 seconds. For an officer which is more experienced or an officer who has had certain training, the mean time can be 3 seconds. An inexperienced officer could have a mean time of 9 seconds. Furthermore, experience can also impact the accuracy of the decision. An inexperienced officer could have a higher probability for false positives and false negatives which directly impacts both security and efficiency.

If the X-ray officer can not clear the bag, the luggage arrives at the bag checker officer. The bag checker manually searches the contents of the bag. If there was a forbidden item on the X-ray image of the luggage, the bag checker aims to find this item and confiscate it. If there was doubt on the X-ray image of the luggage, the bag checker aims to make sure there are no forbidden items in the luggage. Human factors impact the performance of bag checking since officers can have different mean times of checking a bag.

B.2. MODEL DESCRIPTION

The security screening process described in Section B.1 has been implemented by Knol in an agent-based model [63]. This model environment was simplified, as the security checkpoint is the main component of this study.

The environment is a security checkpoint an its area where the objects of the model are a physical queue and a security lane, seen in Figure B.1. Any number of flights with number of passengers can be added to the environment. Two security sensors are part of the environment: the WTMD, and the X-Ray sensor. Both sensors were modelled through signal detection theory. Knol has calibrated the WTMD TPR and FPR to data found in practice where $0.45 < TPR < 0.97$, and $0 < FPR < 0.2$ [63–66]. The calibrated parameters for the WTMD signal are seen in Figure B.2. If the variable threshold is set to 3.090, the security focus is low. The TPR and FPR then correspond to 0.45 and 0.016 respectively [32]. If the variable threshold is set to 0.842, the security focus is high. The TPR and FPR then correspond to 0.97 and 0.2 respectively [32]. As the trade-off between security and efficiency focus has already been explored by Knol, a fixed threshold is set corresponding to a security focus of 0.75. This value was chosen because it represents a realistic and effective value regarding TPR and FPR [32]. Concerning the X-ray sensor, it is assumed sensor provides a perfect image of the scanned luggage.

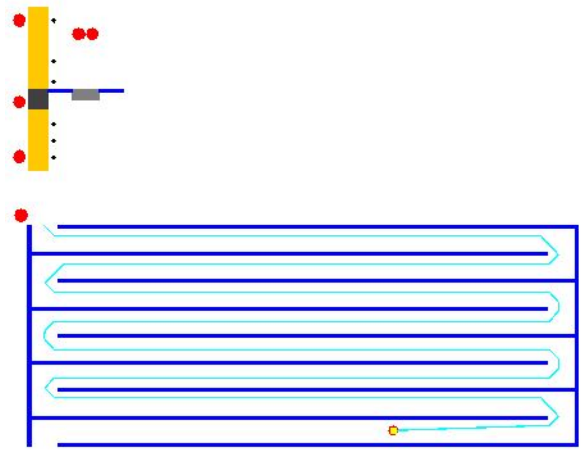


Figure B.1: The security checkpoint environment of the model.

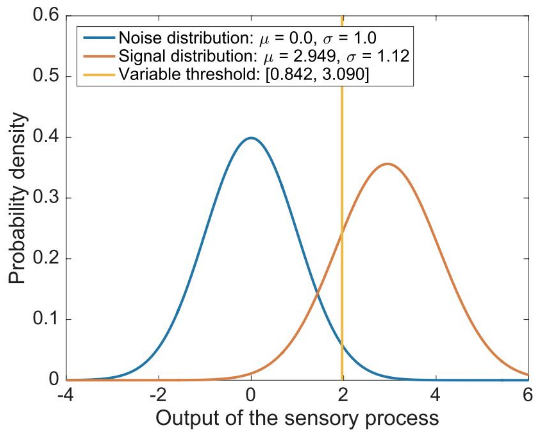


Figure B.2: The noise distribution, signal distribution, and variable threshold for the WTMD.

Two type of agents are defined in the model: operators and passengers. The goal of the passengers is to pass through the security checkpoint as quick as possible. Passengers have carry-on luggage which can contain an illegal item. An illegal item on the passenger body is also an option, but not considered for the experiments in this study. Kirschenbaum states that a leisure flight requires an extra luggage check every

2-3 passengers, whereas for a business flight every 7-9 passengers [10]. This corresponds to a proportion of 0.1270 for business passengers, 0.4167 for leisure passengers, and 0.2067 for average passengers, which carry a forbidden item.

Different security operator types are defined: x-ray, luggage check, and physical check. Each operator makes a decision on whether a forbidden item is present or not. Signal detection theory, used for the sensors, is too limited as it does not take into account human factors. Therefore, the Ratcliff diffusion model was used to give the operators a more cognitive human way of making decisions [67]. The accumulation of evidence for a decision is determined by the drift rate parameter, v, as seen in Figure B.3. A positive drift rate will in most cases hit the upper response boundary, which results in a positive response, thus an illegal item is found. In contrary, a negative drift rate will in most cases hit the lower response boundary, which results in a negative response, thus no illegal item is found. The distance between the upper and lower response boundary is indicated by threshold (a) in Figure B.3. A large a will result in accurate decisions and long decision times. Furthermore, the bias is a fixed distance in which an operator is removed from the response boundaries. Moving the bias closer to the upper response boundary increases the security focus, thus a high TPR and FPR. Moving the bias to the lower response boundary decreases the security focus, thus a low TPR, and a low FPR. Futhermore, the non-decision time, seen in Figure B.3, is the time between presentation of a stimulus, and the start of the decision process. Since the different operators have to make a comparable decision on whether an item is forbidden or not, Knol has made the assumption that the decision times are equal [32]. Thus, the non-decision times make the difference between the total response time of an operator.

In this study, the response time is defined as the decision time plus the non-decision time, seen in Equation B.1.

$$\underset{Z1}{|}\overbrace{\underbrace{\rule{4cm}{0.4pt}}_{\text{Decision Time}}\underset{B}{|}\underbrace{\rule{7cm}{0.4pt}}_{\text{Non-Decision Time}}}^{ResponseTime}\underset{Z2}{|} \tag{B.1}$$

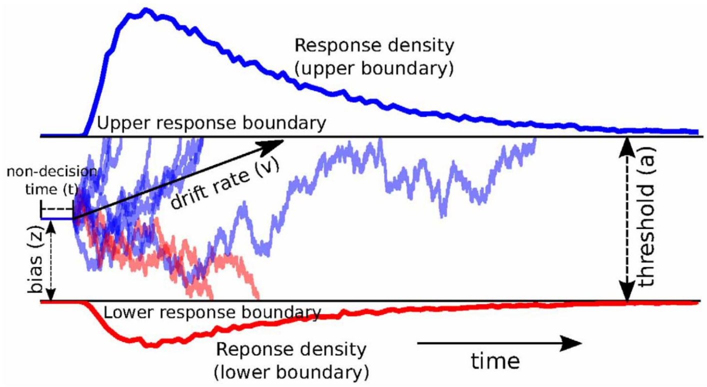


Figure B.3: Graphical representation of the Ratcliff diffusion model. [68]

In addition, McCauley's biomathetmatical fatigue model has been used to account for effects of sleep and sleep loss on operator performance [69]. Walsh has combined McCauley's fatigue model with Ratcliff's diffusion model to relate fatigue with decision making. As time progresses during a day, the fatigue score of operators increases, resulting in a lower absolute drift rate, which implies that decisions take longer and are less accurate. This linear function is $V_{dynamic} = \alpha_v \cdot F(t) + \beta_v$, where α_v is the decrease per fatigue point, $F(t)$ is a function of the fatigue score, and β_v is the intercept of the drift rate.

B.3. MODEL MODIFICATIONS

To incorporate differences in operator performance, several modifications to the model were made. Changing diffusion operator settings in Knol's model changes the performance of all operators at the same time. In reality, each specific operator will have a different speed, accuracy, and fatigue score. If speed is γ, accuracy shift is α, accuracy is σ and ϕ is fatigue, $O_{xr}(\gamma,\alpha,\sigma,\phi)$, $O_{lc}(\gamma,\alpha,\sigma,\phi)$, and $O_{pc}(\gamma,\alpha,\sigma,\phi)$, denote the specific variables for an X-ray operator, luggage check operator, or physical check operator.

SPEED

Regarding speed, Knol makes the assumption in the model that the different operators have to make a comparable yes or no decision regarding a passenger with a forbidden item, and that therefore their decision time is equal [32]. Since this is not the case in reality, a non-decision time is added to the decision time to represent a realistic time it takes for an operator to perform its duty. Knol's total decision times and non-decision times from calibrating the model on RTHA data are seen in Table B.1. The decision times have been found by subtracting the non-decision time from the total decision time. The X-ray decision times are different depending on how many boxes the passenger needs. The assumption Knol makes that the decision times are equal, does not seem to hold by the provided numbers. To see if the model corresponds to the provided decision times, the model has been simulated for 1200 passengers with the corresponding non-decision times. The non-decision times were then set to zero, and the simulation was run again to calculate the simulation result of the non-decision time and decision time. It is observable that the simulation decision times are similar to each other. Since operators are quicker in collecting evidence when an item is allowed versus when an item is forbidden [32], the average simulation X-ray decision-time is, for example, lower than the average simulation decision-time of a luggage check operator. This is because luggage check operators interact mainly with passengers with forbidden items, whereas X-ray operators interact with every passenger. The average X-ray decision time, thus also depends on the ratio of forbidden passengers to non-forbidden passengers. Observing the non-decision times of the simulation, it is seen that they are close to the non-decision times that are used as input. The mismatch ranges from approximately 5 % to 15 %.

Table B.1: Average Decision Times (DT) in seconds, calculated by subtracting non-decision times from the total decision times that were fitted on data from RTHA [32]. The data from the simulation is collected from 1200 simulated passengers.

	Total DT	Non-DT	DT	Non-DT (simulation)	DT (simulation)
Physical Check	43.00	32.71	10.29	29.60	10.40
Luggage Check	104.67	61.82	42.85	57.84	12.07
X-Ray 1 box	10.28	3.67	6.61		
X-Ray 2 box	16.44	7.86	8.58		
X-Ray 3 box	20.82	12.24	8.58		
X-Ray 4 box	21.00	15.36	5.64		
X-Ray Average	15.36	7.42	7.93	6.25	8.41

The introduced speed parameter, γ, is actually a multiplication constant for the non-decision time inputs. A γ equal to 0, will results in a non-decision time of zero, and a γ of 2.0, will result in a twice as large non-decision time as the calibrated non-decision time.

ACCURACY: Z-SHIFT

Regarding accuracy, parameter α was introduced to manipulate the bias, z, to 'shift' the bias during the decision making process. Generally, increasing bias will result in a higher hit rate, and a higher false positive rate [32]. Knol has experimented on how changing 'security focus' impacts the average queue time, where security focus changes the threat threshold and the bias of the diffusion process. The introduced parameter, shifts bias by adding the value of α for passengers with forbidden items, and subtracts α for passengers with allowed items. The result is that an operator with $\alpha < 0$ has a lower hit rate, and a higher false positive rate. Vice versa, an operator with $\alpha > 0$ will have a higher hit rate, and a lower false positive rate.

However, changing bias does not only affect the accuracy. Generally, increasing bias results in shorter decision times for 'positive decisions', and longer decision times for 'negative decisions' [32]. For a positive decision, the evidence reaches the upper response boundary, thus the operator decides that there is a forbidden item. Likewise, for negative decision, the evidence reaches the lower response boundary, thus the operator decides that there is no forbidden item. For parameter α, a positive α will result in shorter decisions for both positive and negative decisions.

With both α and γ, it is possible to express a wide range of operator configurations. If the shorter decision time from setting accuracy α is not desired, the γ can be adjusted to obtain the correct total decision time of the operator.

It is required to translate the α values to TPR and FPR values of the different operators. If data is available on TPR and FPR of a certain operator, then the corresponding α for this operator can be selected. These values are found in Table B.2. It is seen that for the baseline value of $\alpha = 0$, the operator TPRs are around 97.6 %, and the FPRs are close to 1.2 %. This is a too high of a detection performance. Schwaninger uses $d' = z(TPR) - z(FPR) = \frac{\mu_f - \mu a}{\sqrt{\frac{1}{2}(\sigma_f^2 - \sigma_a^2)}}$ for a measure of security performance [70]. Knol uses this measure to find a value of $d' = 2.778$ for the average security performance, where the average security operator has a TPR of 95.0 % and a FPR of 10 % [32]. A more realistic baseline for the operators would therefore be for the value of $\alpha = -0.15$ as the operator TPRs are approximately 95.0 %, whereas the FPRs are approximately 8.0 %. This is a more realistic average security operator performance.

Table B.2: True positive and false positive rates of operators corresponding to different α values in the simulation.

α	-0.2	-0.15	-0.1	-0.05	0	0.05	0.1	0.15	0.2
WTMD + O_{pc} TPR	0.8595	0.873	0.884	0.8919	0.8972	0.9021	0.906	0.9095	0.9137
WTMD FPR	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
O_{xr} TPR	0.9306	0.9487	0.9587	0.967	0.9769	0.9818	0.9864	0.9906	0.9934
O_{xr} FPR	0.1721	0.083	0.0404	0.02137	0.01236	0.0069	0.004182	0.002598	0.001727
O_{lc} TPR	0.9311	0.9504	0.9573	0.9663	0.9756	0.9806	0.9851	0.9853	0.9937
O_{lc} FPR	-	-	-	-	-	-	-	-	-
O_{pc} TPR	0.9328	0.9493	0.9609	0.9693	0.9757	0.9803	0.9849	0.9893	0.9931
O_{pc} FPR	0.1734	0.07954	0.03809	0.01815	0.01155	0.0093	0.00387	0.00141	0.000784

ACCURACY: SECURITY FOCUS

A limitation of the implemented diffusion model in the agent-based model is that it is not possible to directly specify TPR and FPR values of operators. The response times of the three different operators in the agent-based model calibration of a single set of response times. Furthermore, this calibration has no collected data of operator individual TPR and FPR values, which is why ranges of 45-97 % for TPR and 0-20 % for FPR were used obtained from literature [32]. A better implementation of this would require data sets of the response and response times of different individual operators, which were not available in this study. Furthermore, if this was available, it would be useful to integrate a calibration algorithm, such as DMAT [71], directly into the agent-based model such that an individual operator and its diffusion model parameters can be specified directly through its responses and response times. The operator calibration in the model assumes $0.45 < TPR < 0.97$, and $0 < FPR < 0.2$, as provided in literature [63–66], for which a group setting can be set through security focus. This assumption was the reason why the previous introduced parameter, z-shift, was introduced. However, the weakness of this is that there is no direct control over the TPR and FPR. Furthermore, the range of the tested z-shifts seen in Table B.2, does not include possible TPR values below 0.93. If the z-shift would be further decreased, this would give TPRs < 0.93, but it would give FPRs > 0.20, which is unrealistic. Therefore, to explore the complete realistic range of TPRs, security focus, used in Knol's study as a group parameter [63], was implemented as an individual operator specification. The introduced parameter, σ, set the TPR of the operator. Table B.3 shows TPRs and FPRs of varying SF focus values. SF values from 0.15 to 0.85 are considered due to the using $\alpha = -0.15$ as a correcting factor.

Table B.3: True positive and false positive rates of operators corresponding to different security focus values.

Security Focus	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85
WTMD + O_{pc} TPR	0.0706	0.0734	0.0739	0.0668	0.0690	0.0623	0.0572	0.0603
WTMD FPR	0.2393	0.5295	0.6618	0.7612	0.8118	0.8250	0.8602	0.8590
O_{xr} TPR	0.0009	0.0026	0.0040	0.0069	0.0134	0.0301	0.0790	0.2434
O_{xr} FPR	0.2887	0.5731	0.7448	0.8243	0.8873	0.9159	0.9490	0.9581
O_{lc} TPR	0.2969	0.5670	0.7182	0.8460	0.9000	0.9209	0.9501	0.9625
O_{lc} FPR	-	-	-	-	-	-	-	-
O_{pc} TPR	0.2679	0.5895	0.7298	0.8293	0.8910	0.9133	0.9497	0.9608
O_{pc} FPR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0361	-	0.2242

FATIGUE

Knol's fatigue analysis was performed with a single fatigue score curve for the average person [63]. To expand the model, fatigue curves of a person sleeping 4 hours, 6 hours, 8 hours were implemented. The implementation makes use of the bio-mathematical model equations established by McCauley [69]. It is similar to the fatigue score predictions of Walsh [72]. The fatigue score after 2 weeks of constant sustained sleeping schedule was used. The scores are seen in Figure B.4.

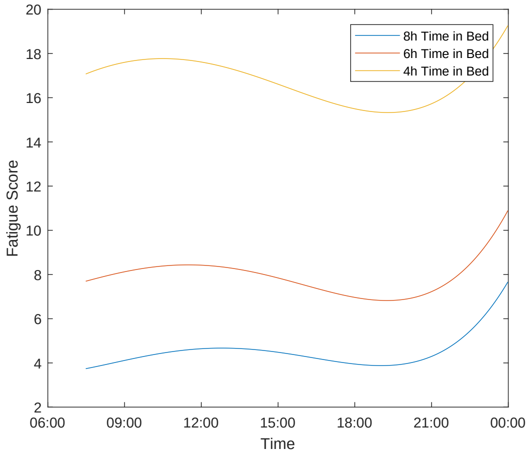


Figure B.4: The fatigue score curves of people that sleep 4 hours, 6 hours, and 8 hours a day, sustained for two weeks.

RECALIBRATION

As mentioned in Appendix B and Section B.3 , the agent-based model was calibrated to $0.45 < TPR < 0.97$, and $0 < FPR < 0.2$ [63–66]. Trying to replicate TPR and FPR results by varying security focus [32], mismatches were found in which TPR is overestimated and FPR is underestimated as seen in Table B.4 which shows TPR and FPR for an x-ray operator with the 'baseline calibration'.

Table B.4: True positive and false positive rates of an x-ray operator corresponding to different security focus values.

Security Focus	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1
O_{xr} TPR	0.000174	0.000241	0.000528	0.000964	0.001854	0.004845	0.01236	0.0357	0.1301
O_{xr} FPR	0.4706	0.7162	0.8351	0.9026	0.9364	0.9602	0.9769	0.9851	0.9919

The correct procedure would be to fully re-calibrate all diffusion related parameters, which was not possible due to unavailability of data on responses and response times. Therefore, Knol's calibration was used a baseline calibration, and the accuracy shift parameter introduced in Section B.3 was used to fix an approximate average operator at an approximate average security focus. A security focus of 0.75 is considered an approximate average security focus as discussed in Section B.2. An empirical study conducted found $TPR = 0.95$ and $FPR = 0.10$ for the average operator [33]. Thus, as seen in Table B.2, setting $\alpha = -0.15$ results in an average TPR of 0.95 with a average FPR of 0.08, which is why $\alpha = -0.15$ is used for a baseline operator.

Furthermore, similar problems were encountered as in Knol's study [32], in which the throughput is lower than than a realistic throughput. Reasons for this could be that in reality the operation of dropping and picking up luggage, and moving through the security checkpoint, occurs more dynamic than in the model. For example, in the model 3 passengers at the time can drop or pick up their luggage while the rest of the passengers has no activity. Knol's solution to this is a scaling factor, the throughput adjustment factor, which scales the non-decision times. Setting a low throughput adjustment factor therefore increases the throughput. A throughput adjustment factor of 0.65 matched the expert-advised goal throughput of 2.6, in Knol's study. Since then, a new data-driven analysis of Rotterdam The Hague airport was conducted, where in reality the average throughput was found to be close to 2.0. In the current model, if the throughput adjustment factor is set to 0.65, the found throughput is approximately 1.4. Therefore, additional measures have to be taken to calibrate the model to reality.

While making scaling adjustments to the model, it is also possible to adjust the luggage drop and collect times, and also the decision time. A throughput of 2.0 was found when the throughput adjustment factor is set to 0.60, the luggage correction factor is set to 0.6, and the decision time factor is equal to 0.50. Besides increasing the throughput, the decision time factor was set to 0.50 to make lower response times possible, since if this is unchanged, the average decision time of an X-ray operator will result in 8.4 s. This is too high, considering that Schwaninger finds average response times as low as 3.5s [12]. The 3.5s was achieved through specialised training, however values of 5s [12] to 7s [73], are more realistic noted average response times. The decision time factor of 0.50 reduced the average decision time of an X-ray operator to 4.5 s. Thus, if the non-decision time is set to zero, the minimum average response will be 4.5 s.

C

METHOD: MULTI-CRITERIA DECISION MAKING

This section elaborates on the normalization procedure that is used in the MCDM method. The motivation for using the simple additive weighting for aggregation and discrete choice analysis for obtaining criteria weights is given in Part I. The theory of discrete choice analysis and its application towards the study is provided Appendix D.

C.1. NORMALIZATION

Since the criteria have different units, it is required to perform normalization to make the data dimensionless and obtain a single score. Vafaei et al. reviews five common normalization techniques and compares their performance on a case study using AHP [74]. Table C.1 shows the 5 common normalization techniques that are used in that study.

Table C.2 shows an example of applying the first four normalization techniques on three alternatives with different performance on efficiency (queue time), and security (vulnerability). Logarithmic normalization is more uncommon and was therefore not applied in the example. The values in the table are V_i as seen in the weighted sum model aggregation shown in Equation C.1.

$$V_i = \sum_{n=1}^{n} w_j q_{ij} \tag{C.1}$$

First the normalization is applied per criterion by which q_{ij} is obtained as seen in Equation C.1, where i is an alternative and j is a criterion. Then the obtained score is multiplied by the respective criterion weight w_j. Since queue time, vulnerability, and cost are all cost criteria, the cost formulas of the normalization techniques were used. The weights in this example were set such that 200 seconds of queue time is equal to 0.01 hit rate. The weights are then found by applying $w_j = \frac{u_j/v_j}{\sum_{j=1}^{n} u_j/v_j}$, where u_j is the unit for criterion j, and v_j is value input such as $200s$ and 0.01 hit rate. v_j is in this case defined beforehand, but is normally found by $v_j = \frac{u_j}{w_j}$. The units of measurements are $u_1 = 60$ for efficiency, $u_2 = 0.01$ for security, and $u_3 = 1$ for cost. The obtained weights found are then 0.23 for efficiency, and 0.77 for security.

Considering the example in Equation C.1, the three alternatives should be valued equally due to 200 seconds of queue time being valued 0.01 hit rate. However, it is seen that the different normalization techniques give different results. Technique $N1$ values alternative 1 as the best, for $N2$ all alternatives are valued equal, for $N3$ alternative 3 is valued the best, and for $N4$ alternative 1 is valued the best. Thus, the linear max-min technique is the only technique which is reliable for the described situation.

However, the linear max-min technique was not found to handle all situations as expected. For example, if we increase the queue time of alternative 3 by 50 seconds compared to Table C.2, seen in Table C.3, it is seen that alternative 2 is valued the highest. However, considering the values, alternative 1 and 2 should be valued equally. This score judgement occurs due to the unit information the weights hold. This was corrected by introducing normalization technique $N6$ for which the cost criteria is defined Equation C.1. As seen by the values of V_i in Table C.3, alternative 1 and 2 do have an equal performance.

$q_{ij} = \frac{r_{max}-r_{ij}}{u_j}$ (C.2)

.

Moreover the score of using normalization $N6$ has a meaning as it can be interpreted more than just a number. For example, the difference between alternative 1 and alternative 3 is equal to 0.192. Considering the 50 seconds increase of queue time of alternative 3, this is equal to $50/v_j = 50/(\frac{u_j}{w_j}) = 50/(\frac{60}{0.23}) = 0.0192$. Thus, v_j is a constant property for which all v_j are equally valued. If the difference between best and worst alternative is equal to 1.0, considering only efficiency and security, and $v_1 = 300$ & $v_2 = 0.015$. Then, if the difference between alternatives is only on queue time then this is equal to 300s. If the differences between alternatives is only in hit rate, then this is equal to 0.015. If the differences are due to a combination of criteria, due to v_j, the value of 1.0 gives a decision maker an idea of large the benefit is by choosing the best alternative over the worst.

Table C.1: Normalization techniques [74].

Normalization Technique	Condition of Use	Formula
Linear: Max (N1)	Benefit criteria	$q_{ij} = \frac{r_{ij}}{r_{max}}$
	Cost criteria	$q_{ij} = 1 - \frac{r_{ij}}{r_{max}}$
Linear: Max-Min (N2)	Benefit criteria	$q_{ij} = \frac{r_{ij}-r_{min}}{r_{max}-r_{min}}$
	Cost criteria	$q_{ij} = \frac{r_{max}-r_{ij}}{r_{max}-r_{min}}$
Linear: sum (N3)	Benefit criteria	$q_{ij} = \frac{r_{ij}}{\sum_{i=1}^{m} r_{ij}}$
	Cost criteria	$q_{ij} = \frac{1/r_{ij}}{\sum_{i=1}^{m} 1/r_{ij}}$
Vector normalization (N4)	Benefit criteria	$q_{ij} = \frac{r_{ij}}{\sqrt{\sum_{i=1}^{m} r_{ij}^2}}$
	Cost criteria	$q_{ij} = 1 - \frac{r_{ij}}{\sqrt{\sum_{i=1}^{m} r_{ij}^2}}$
Logarithmic normalization (N5)	Benefit criteria	$q_{ij} = \frac{ln(r_{ij})}{ln(\prod_{i=1}^{m} r_{ij})}$
	Cost criteria	$q_{ij} = \frac{1-\frac{ln(r_{ij})}{ln(\prod_{i=1}^{m} r_{ij})}}{m-1}$

Table C.2: Example of MCDM input data. The last four columns represent V_i of the alternatives by applying normalization techniques N1, N2, N3, and N4 as given in Table C.1 .

Normalization Technique	Efficiency	Security	Cost	N1	N2	N3	N4
Alternative 1	500	0.08	14	0.093567	0.210526	0.405467	0.198527
Alternative 2	700	0.07	14	0.073099	0.210526	0.460712	0.181955
Alternative 3	900	0.06	14	0.052632	0.210526	0.535762	0.165382

Table C.3: Example of MCDM input data. The last two columns represent V_i of the alternatives by applying normalization techniques N2, and N6, a new normalization technique.

Normalization Technique	Efficiency	Security	Cost	N2	N6
Alternative 1	500	0.08	14	0.230769	1.730769
Alternative 2	700	0.07	14	0.24359	1.730769
Alternative 3	950	0.06	14	0.230769	1.538462

D

METHOD: DISCRETE CHOICE ANALYSIS

To obtain criteria weights for MCDM problems, classic MCDM methods are often used. Discrete Choice Analysis (DCA) is an alternative method of determining weights which makes use of a Discrete Choice Experiment (DCE) to study the effects of attribute levels on the stated preference. DCA is almost never mentioned next to MCDM in scientific literature. This could be because MCDM methods consist of a complete methodology where several steps are required, whereas DCA is a single step that allows to obtain weights. An MCDM method which uses DCEs as part of its methodology does not exist. A reason for this could be that DCA is a separate field within scientific literature with many other applications. However, there are examples of literature in which a DCE is used to determine criteria weights for MCDM [75], or is considered as an option next to classic MCDM weighting methods [76].

The main motivation for using DCA for weight estimation is that many MCDM methods use tradeoffs and pariwise comparison which require experts to value criteria and alternatives. This valuation leads to inaccurate responses since evidence from cognitive psychology indicates that people cannot accurately report on why they make certain decisions [77]. It is also observed that people cannot directly observe their cognitive process, but are able to make good choices between alternatives [77]. Furthermore, the field of DCA has a cognitive psychological basis. Therefore, using a DCE for weight estimation allows to obtain expert information more accurate compared to the weight estimation from MCDM methods.

D.1. DCE DESIGN THEORY

DCA involves designing a DCE to study the effects of attribute levels on the stated preference [78]. The general framework of the DCA includes the following components [79]:

1. Decision-maker: an individual or group of individual which are used for the choice modelling experiment.
2. Alternatives: the different options that are available to the decision-maker.
3. Attributes: the variables that specify an alternative. The set of alternatives is known as the choice set.
4. Decision rule: the process by which a decision-maker selects an alternative.

The design of the DCE consists of five stages seen in Figure D.1

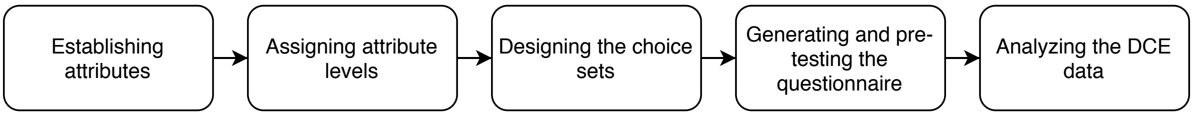


Figure D.1: The five stages of designing a DCE [78].

The first step consists of establishing attributes. From a MCDM viewpoint, the attributes are the criteria. For the investigated case study, security, efficiency, and cost are the chosen airport terminal attributes that make up an alternative.

The next step it to assign attribute levels. Levels should be chosen which are realistic and meaningful such that they represent reality as accurate as possible. Often, two or three attribute levels are specified, since too many attribute levels can complicate the choice process.

The third step involves designing the choice sets. Common experimental designs are orthogonal, full factorial, orthogonal fractional factorial, and efficient designs. The user manual of Ngene [1], an extensive software package used for designing DCEs, describes the theory behind these designs and their properties, advantages, and disadvantages. A full factorial design allows estimation of main effects and interaction effects, but is often unpractical due to the large number of choice sets that is obtained. Therefore, fractional factorial designs are used, which reduce the number of choice sets drastically, but are not able to estimate interaction effects. The aim of a fractional factorial design is to be both orthogonal and balanced [78]. The attributes of an orthogonal fractional factorial design are statistically independent. For a balanced design, the attribute levels occur equally often such that the variance of the parameter estimates are minimized. It is also required to decide what the decision rule the respondents are going to follow. Most discrete choice models use the decision rule based on utility theory [79, 80]. This decision rule follows that the decision-maker chooses the alternative which brings the most value, thus maximizing utility. An alternative decision rule is based on regret theory, for which the aim of the decision-maker is to minimize regret.

The fourth step involves generating the DCE and pre-testing it. A decision has to be made on how much choice sets to present to each respondent. It was shown that you can ask 20 choice sets before fatigue can degrade the data quality [81]. Generally, 18 choice sets is considered a practical limit until boredom sets in [78]. Conducted studies most often used 9 to 16 choice sets [82]. The DCE has to be presented clearly to the respondent, and have an introduction describing the attributes, the situation, and the goal. This also indicates if the respondents have to give their stated preference, or revealed preference. Pre-testing can provide insight whether the DCE attributes and levels are selected correctly [78].

The final step is to analyze the DCE data. The final DCE is sent out to respondents such that data can be collected. The analysis of DCEs involves the use of discrete choice models which are regression models. Discrete choice models statistically relate the made choices with the attributes of the of the alternatives. Several types of discrete choice models are logit models, GEV models, probit models, and mixed logit models. An extensive overview of the theory behind the models is provided by Train [83]. Logit models are by far mostly used as discrete choice model [83]. It rests on the assumption that unobserved factors are uncorrelated over the alternatives, and have the same variance for all alternatives. Furthermore, the assumption is made that each choice is independent of the others. GEV models, probit models, and mixed logit models, do not assume independence of unobserved factors.

D.2. DCE DESIGN

The DCE design process of Section D.1 was followed to establish the DCE for the case study. First a pilot study was conducted to see if the DCE was designed properly.

The selected attributes are the airport terminal performance areas of security and efficiency, and cost. Security is measured by hit rate, efficiency by average queue time, and cost by average cost of the operators. Since the dependency effects between the attributes are not significant, the assumption is made that the attributes are independent.

The selected attribute levels for security and efficiency were based on scenario tests performed in the agent-based model, whereas cost was initially based on operator salaries. The attribute levels for the pilot study are seen in Table D.1.

Table D.1: The selected attribute levels of the DCE designs.

	Pilot DCE	Final DCE
Efficiency	[3 5 7]	[7 12 17]
Security	[91 93 95]	[85 90 95]
Cost	[12 14 16]	[30 35 40]

Efficient designs based on the utility maximization decision rule were chosen as the experimental design for the choice sets. The utility function of Random Utility Maximization (RUM) is represented by a linear-additive function seen in Equation D.1.

[1] http://www.choice-metrics.com/NgeneManual120.pdf, retrieved on 02/10/2019

$$U_{in} = V_{in} + \epsilon_{in} = \sum_k \beta_k \cdot x_{ink} + \epsilon_i \tag{D.1}$$

Here U is utility, V is observed utility, ϵ is unobserved utility, i is alternative, n is an individual decision maker, β_k is the estimated parameter (regression coefficient) of the kth attribute, x is the attribute level.

The aim of efficient designs is to optimize use of data such that statistically significant parameter estimates can be produced for small sample sizes [84]. Since a small sample size is expected, efficient designs may result in obtaining a statistically significant parameter. Efficient designs use prior parameters such that the design algorithm knows where the trade-off between the parameters is approximately to happen. Priors are generally obtained by conducting a pilot study.

The design of the choice sets and the DCE was performed using an software tool named Robust Design Generator (RDG) [85]. This tool allows generating DCEs with an efficient design. It minimizes the D-error, which is the most common statistic efficiency measure for experimental designs [85]. A number of 10 choice tasks was selected as input. As there is no information regarding the size of the priors, it is recommended to use a very small value if it is known whether the the parameter is positive or negative. Thus, 0.001 was selected for security since hit rate is a benefit criterion, whereas -0.001 was used for efficiency and cost as they are cost criteria. The model is optimized for the RUM decision rule. The experimental design is unlabelled with three alternatives per choice task.

Table D.2 shows the design of the pilot DCE. The second row indicates the first choice task. X12 indicates alternative 1 of attribute 2. Attribute 1 is efficiency, attribute 2 is security, and attribute 3 is cost. The DCE was conducted through Microsoft Forms.

Table D.2: The design of the pilot DCE.

Task	X11	X12	X13	X21	X22	X23	X31	X32	X33
1	3	91	12	3	95	16	7	93	14
2	3	91	12	5	93	16	7	95	12
3	3	91	12	5	93	16	7	95	16
4	3	91	16	5	91	12	7	95	12
5	3	91	16	5	93	12	7	95	16
6	3	91	16	5	95	14	7	91	12
7	3	91	16	5	95	16	7	95	12
8	3	93	14	7	91	12	7	95	16
9	3	95	16	7	91	12	7	93	14
10	3	95	16	7	91	12	7	95	14

The pilot study was analyzed through a package of Matlab scripts that allow for estimation of models for DCEs [2]. The frequently used Multinomial Logit model was used, which assumes independence of unobserved factor. The model is used to predict the probabilities of the possible outcomes of categorical dependent variables, given a set independent variables. The dependent variables to be estimated are the regression coefficients, β_k, as seen in Equation D.1, wheres the independent variables are x_{ink}. The regression coefficients are estimated using maximum likelihood estimation for fitting the model to the data. The probabilities are calculated by Equation D.2.

$$P_i = \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}} \tag{D.2}$$

Here P_i is the probability for alternative i, V_i is observed utility for alternative i, and $j = 1, ..., J$ is the set of alternatives. A total of 3 respondents were collected for the pilot study. The results of the parameter coefficients were all statistically insignificant. The primary reason was the design of the DCE, since the respondents did not participate in trading off the attributes. Since the queuing times were all below 10 minutes, the respondents maximized hit rate for all choice sets. Thus no reliable coefficients were found for the attributes. Therefore, new attributes were selected in cooperation with a security expert which are used for the final DCE design seen in Table D.2. Notable changes are increased distance between the attribute levels such that alternatives can be better distinguished by respondents. The attributes of cost were also changed from salaries of

[2] https://github.com/czaj/DCE, retrieved on 03/10/2019

security operators, to the hourly fee the airport pays the security company per operator. Priors from the pilot design were not used for the final design due to statistical insignificance. The design of the final DCE is seen in Table D.3

Table D.3: The design of the final DCE.

Task	X11	X12	X13	X21	X22	X23	X31	X32	X33
1	7	85	30	7	90	40	17	95	40
2	7	85	30	12	90	40	17	95	30
3	7	85	30	17	90	30	17	95	40
4	7	85	35	7	95	40	17	85	30
5	7	85	40	12	90	30	17	95	40
6	7	85	40	12	90	40	17	95	30
7	7	85	40	12	95	35	17	85	30
8	7	90	35	17	85	30	17	95	40
9	7	95	40	12	85	35	17	95	30
10	7	95	40	17	85	30	17	95	35

D.3. FINAL DCE RESULTS

The final DCE was filled in by a total of 7 respondents. The respondents are security experts with the functions security advisor, duty manager, and security specialist. A total of 69 observations were made since one respondent did not complete one choice set. The resulting regression coefficients are seen Table D.4.

Table D.4: The regression coefficients, standard error, and p-value, estimated by Multinomial Logit regression.

Variable	Coefficient	Standard error	p-value
1: Efficiency	-0.1388	0.0457	0.0024
2: Security	0.2989	0.0583	0.00004
3: Cost	-0.0315	0.0366	0.3893

It is seen that the security coefficient is largest in absolute terms. This implies that security has the largest importance compared to efficiency and cost. The standard error is relatively high compared to the coefficient values. The p-value of efficiency and security are below 0.05 indicating that these coefficients are statistically significant, whereas the cost is not. Respondents hardly felt the need to trade-off with cost.

Mcfadden's pseudo-R^2, a goodness-of-fit measure, is equal to 0.3313. Values between 0.2 and 0.4 for Mcfadden's pseudo-R^2 indicate a very good fit of the regression model to the data [86].

The process of obtaining the weights of the attributes is seen in Table D.5. , and follows the process as seen in [87]. The part-worth utility is obtained by multiplying the obtained regression coefficients with the attribute levels. The utility range indicates the difference between the minimum and maximum part-worth utility in absolute terms. The weight is then found by dividing the attribute utility range with the sum of all utility ranges.

Table D.5: Parameters found in the process of estimating the weight (attribute importance) of the variables.

Attribute	Coefficient	Attribute Levels	Part-Worth Utility	Utility Range	Weight
1: Efficiency	-0.1388	[7 12 17]	[-0.97 -1.67 -2.36]	1.39	0.2958
2: Security	0.2989	[85 90 95]	[25.41 26.90 28.40]	2.99	0.6370
3: Cost	-0.0315	[30 35 40]	[-0.94 -1.10 -1.26]	0.315	0.0671

Ratios of regression coefficients are marginal rates of substitution. This is the rate at which one attribute can be given up for another while maintaining the same utility. Thus, $\frac{\beta_2}{\beta_1} = \frac{0.2989}{-0.1388} = -2.15$, implies that an hit rate unit was found to be the equivalent of 2.15 efficiency units, which means that 0.01 hit rate is valued 129 seconds of queuing time. This is an important measure to determine which alternatives are better compared to others.

E

SENSITIVITY ANALYSIS: MODEL

A fundamental step to understand the behaviour of the complex agent-based system is sensitivity analysis. The main purpose of this sensitivity analysis is to have an increased understanding of how the input and output variables affect each other. Several input parameters were introduced in Appendix B to be able to vary performance of different operators. Speed γ, accuracy shift α, accuracy σ, and fatigue ϕ, of the X-ray operator O_{xr}, luggage check operator O_{lc}, and physical check operator O_{pc} will be varied to determine the impact on the output variables queue time, and hit rate. The sensitivity analysis will reveal which input parameters influence the output the most, and under what conditions.

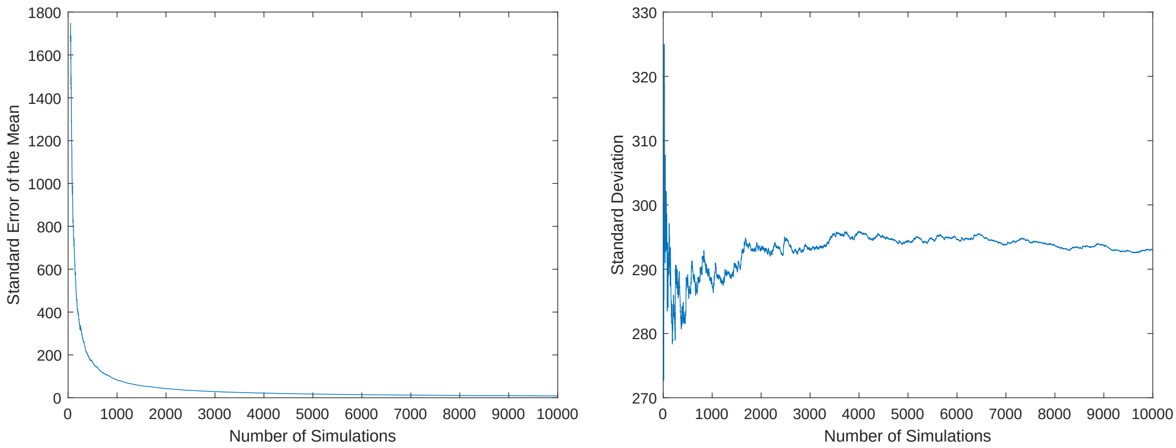
E.1. BASELINE SETUP

The baseline setup involves a single departing flight of 180 passengers. A single security lane is open with one x-ray operators, one luggage check operator, and one physical check operator. This will results in an average queuing time above 10 minutes. Most of the calibrated parameters of Knol's model calibration to Rotterdam The Hague Airport (RTHA), were used for the baseline setup [32, 63]. The arrival distribution was set to $(0.2, 0.8, 0.2)$, which means 20 % of passengers arrive 1:40:00-1:20:00 before departure, 80 % arrives 1:20:00-0:40:00 before departure, and 20 % arrives 0:40:00-0:20:00 before departure. An average forbidden ratio of a charter flight was used, 0.4167 [10]. The correction factors discussed the re-calibration section of Section B.3 were applied to achieve a close throughput to reality. Z-shift of baseline operators was kept constant at -0.15. Furthermore, the security focus is held constant with a value of 0.75, which corresponds to a 75% security focus. The reason for keeping this value 0.75, is because it is expected that airports try to keep their security checkpoint balanced with respect to security and efficiency. Knol finds that the security focus area around 0.75 seems to give good security while keeping processing time low [63]. The operators in the baseline setup have a fatigue score curve corresponding to 8 hours of sleep. The accuracy analysis concerns data of average queue times since hit rate needs less runs for the same accuracy.

E.2. NUMBER OF SIMULATION RUNS

The next step involved determining the minimum number of simulation runs. Increasing the simulation runs decreases the change of the mean and variance. If the simulations are not computationally expensive, it is possible to keep increasing simulations until the output has a satisfactory level of convergence or stability. Ritter analyzes the Standard Error of the Mean (SEM) to determine the number of simulations to run [88]. Since the SEM indicates that the true mean has a 95 % chance of being within range of the estimated mean $\pm 1.96 * SEM$, Ritter suggests to run the model until the estimated range of the mean is satisfactory.

The baseline setup has been run for a number of 10000 simulations to reveal what happens to the SEM and to see how the standard deviation (SD) stabilizes. Figure E.1a shows how the standard error of the mean changes with increasing number of simulations for the average queuing time. From 0 to 1000 number of simulations, the SEM decreases rapidly, after which the magnitude decreases. Figure E.1b shows how the SD of the average queue time behaves with increasing number of simulations. It shows that the SD slightly fluctuates between the 0 to 2000 number of simulations region after which is stabilizes. Examining both these graphs, running 2000 simulations would give a satisfactory low SEM, below 50, and avoid the more turbulent region of the SD.



(a) The Standard Error of the Mean plotted against the number of simulations. (b) The Standard Deviation plotted against the number of simulations.

Figure E.1: Standard Error of the Mean and Standard Deviation plotted against the number of simulations.

Another approach in determining number of simulations is to use the coefficient of variation, c_v, as suggested by Lorscheid [89]. It is equal to the SD, σ, divided by the mean, μ, thus, $c_v = \frac{\sigma}{\mu}$. The stability of c_v is used as a stopping criterion to determine the number of runs. If c_v is plotted against number of simulations for the 10000 runs of the baseline setup, seen in Figure E.2, it is observable that it has strong similarities with Figure E.1b. Furthermore, it is seen that a 'stability of 0.01' would approximately be achieved with a minimum number of 500 to 1000 simulations, since after 500 runs the c_v values are between 0.265 and 0.275. To achieve a stability close to 0.001, it would be required to run a minimum of 2000 simulations. This is because from 2000 to 10000 simulations, the c_v does not exceed 0.2747 ± 0.001

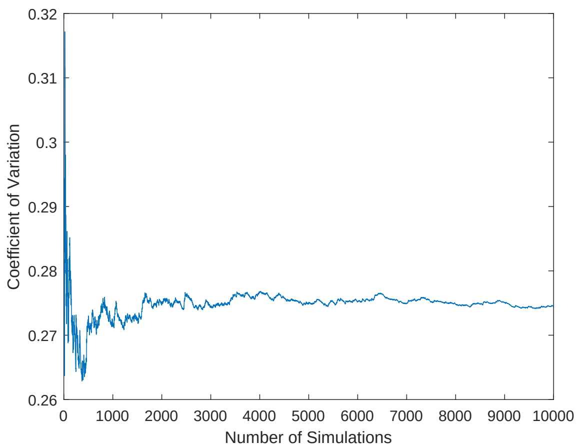


Figure E.2: The Coefficient of Variation plotted against the number of simulations.

Considering figures Figure E.1, and Figure E.2, simulating the baseline setup 2000 times will result in avoiding the less stable region between 0 to 2000 simulations, and therefore give accurate results. Furthermore, for the baseline setup running it 2000 times is not computationally expensive as it includes a single flight. An accurate insight into the parameters will provide more accurate predictions with different flight configurations.

E.3. ONE-AT-A-TIME METHOD

The One-at-a-time (OAT) sensitivity analysis method is one of the most basic methods to analyze how the input parameters affect the output. A baseline configuration is set, after which a single parameter is varied.

The advantage of this method is that the observed output can directly be attributed to the change of the parameter. The disadvantage is that interactions between parameters are neglected. Thus, only first-order effects are obtained.

SPEED

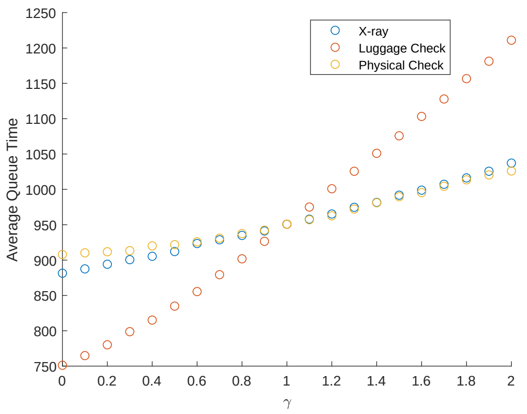
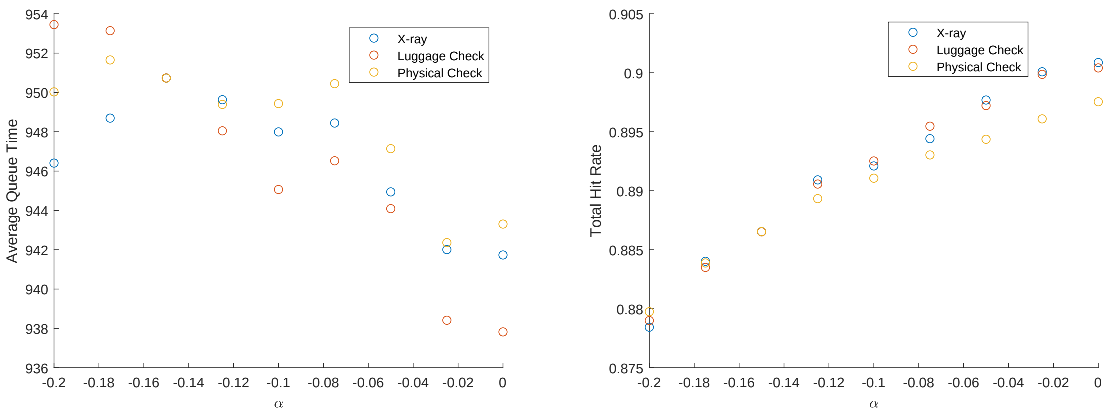


Figure E.3: The average queue time result of the speed parameter value for the different operators.

Figure E.3 shows the average queue time with respect to γ for the three different operators. Besides γ, the throughput adjustment factor also multiplies the non-decision times which are noted in Table B.1. Thus, for example, for a γ of 0.8, the input non-decision time of a physical check operator is $32.71 * 0.65 * 0.8 = 17.0s$.

It is seen the change of γ for the different operators follows a linear trend with the average queue time. The luggage check operator impacts the output the most. whereas the physical check operators impacts the output the least. A change in the trend occurs for the luggage check operators below $\gamma = 0.4$. However, the region of $\gamma < 0.4$ is not of interest for the luggage check and physical check operator as this would decrease the average response time below realistic values. For the X-ray operator this is not the case, because with $\gamma = 0$, the response time will still be around 8 seconds due to the large decision time. Furthermore, it is to be noted that if 2 lanes would be open in the simulation instead of a single lane, the physical check operator influence will be even less as there is a single physical check operator for every 2 lanes in the model.

ACCURACY SHIFT



(a) Average queue time plotted against the Z-shift.

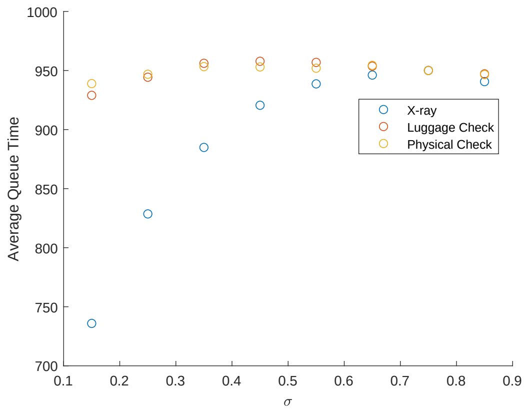
(b) Total hit rate plotted against the Z-shift.

Figure E.4: Z-shift impact on efficiency and security.

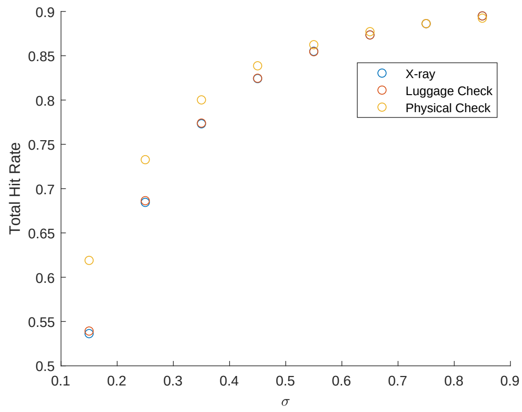
Figure E.4 show the average queue time and the total hit rate plotted against the α. Figure E.4a shows that changing α impacts the average queue time minimally. If a change in average queue time is unwanted due to

α, the γ can be adjusted to set the accuracy and speed to desired values. However, changes in queue-time due to z-shifts are quite small and can be considered negligible. Figure E.4b shows the impact on the total hit rate due to z-shift change of a single operator. As discussed in Section B.3, since z-shift does not cover wide range of TPR values compared security focus, OAT of the z-shift is not specifically interesting as different accuracy of operators will be specified through security focus specification, which is discussed next.

ACCURACY



(a) Average queue time plotted against security focus.



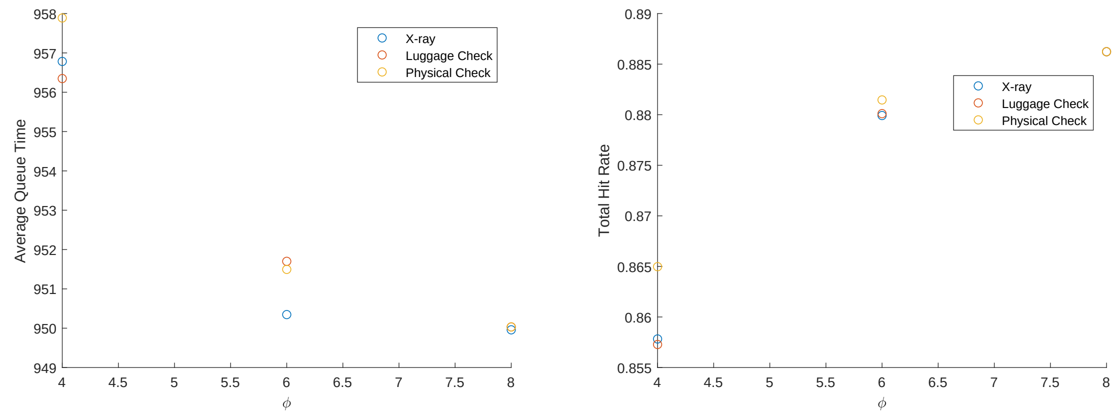
(b) Total hit rate plotted against the security focus.

Figure E.5: Security focus impact on efficiency and security.

Figure E.5 shows the impact of changing the security focus impact on efficiency and security. Figure E.5a shows that average queue time has the largest impact for the x-ray operator compared to luggage check and physical check operators. This indicates that the x-ray operator is the bottleneck of the process in terms of queuing time. If an x-ray operator has a close to zero FPR, then passengers will barely end up at the luggage check operator, which is what saves most time. Figure E.5b shows the curves of the impact on total hit rate with varying security focus. It is seen that x-ray operator and luggage check operator have the same impact on the total hit rate. This is because if an x-ray operator encounters a forbidden item, it is up to the luggage check operator to find the forbidden item, which requires input from both operators.

FATIGUE

Figure E.6 shows the output results of operators with different fatigue score curves. Changes in the output regarding average queuing time for are negligible for all hours of sleep. The decrease on total hit rate with a single operator sleeping 4 hours is approximately 0.03 for all three operators. The difference is slightly less for the physical check operator due to less passenger interaction. Since approximately 7.8 % of the population sleeps less or equal than 5 hours on average [90], the impact of fatigue is not necessarily insignificant.



(a) Average queue time plotted against hours of sleep.

(b) Total hit rate plotted against hours of sleep.

Figure E.6: Fatigue impact on efficiency and security.

E.4. GLOBAL SENSITIVITY ANALYSIS

Since OAT is a local sensitivity analysis method, it does not take into account possible interaction between parameters. Therefore, the next step in the sensitivity analysis would be to use a global method to explore interaction between parameters. Iooss gives a review of the different methods and variations that exist for global sensitivity analysis [91]. Furthermore, Lee states some approaches and challenges of global sensitivity analysis techniques in context of agent-based models [92]. The two notable global sensitivity analysis methods that do no assume linearity are the Morris method, or elementary effects method, and the variance-based sensitivity analysis method, also known as the Sobol method, or Sobol indices.

The Morris method is a screening technique, which means it is based on a discretization of the input parameters [91]. It expands on the OAT method by varying multiple parameters at the time. An advantage of the Morris method is that it allows for computationally efficient experiments, but the method can not quantify the relative importance between factors [93]. If the Morris method would be applied with the discretization used at the OAT method, there would be $(21*9*3)^3 = 182{,}284{,}263$ nodes to explore, which is infeasible. Therefore, it would be required to reduce the exploration space significantly.

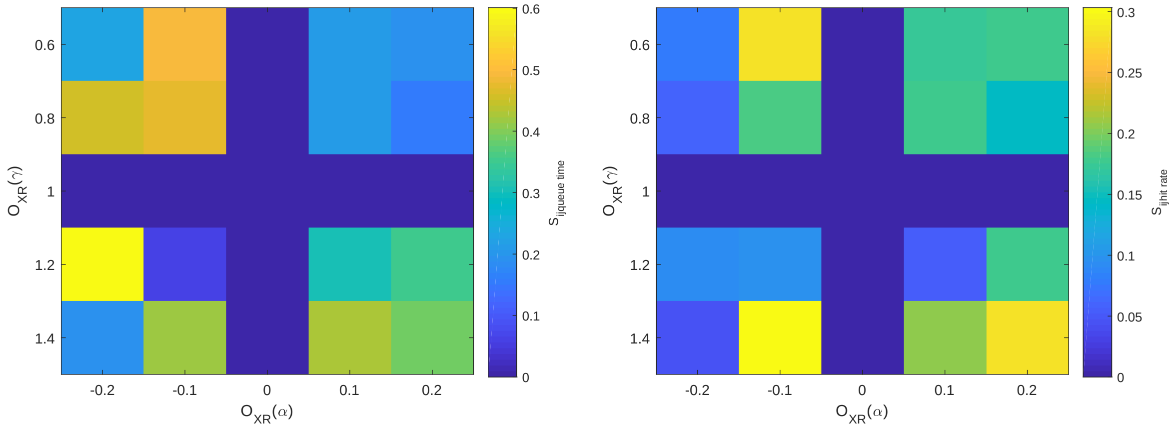
The Sobol method does allow for quantification between the relative importance between factors. This is done by decomposing the variance of the output, and attributing it to the input variances and their interactions. Interactions (second-order effects) are more difficult to observe than first-order effects, but can be well described by variance [93]. For orthogonal factors X_i and X_j, the second order effect is denoted by $V_{ij} = V(E(Y|X_i, X_j)) - V(E(Y|X_i)) - V(E(Y|X_j))$. Thus, V_{ij} is equal to the joint effect, minus the first-order effect of factor X_i, minus the first-order effect of factor X_i.

Furthermore, the sensitivity index is defined as $S_{i_1,\ldots,i_m} = V_{i_1,\ldots,i_m}/V(Y)$ [93]. For the orthogonal inputs of an additive model, $\sum_{i=1}^{k} S_i = 1$ or $S_i + S_j = 1$ as the S_{ij} term is zero in an additive model. More information on the non-additive part of the model can be found through computing the total effects terms S_{Ti} and S_{Tj} [93]. If there is a large difference between, for example, S_{Ti} and S_i, then there is large interaction for that factor.

The GSA was performed for the 15 different second-order effects, for which two are displayed in Figure E.7, and Figure E.8. γ was discretized from 0.6 to 1.4 with steps of 0.2, whereas α was discretized from -0.2 to 0.2 with steps of 0.1. No visual patterns were noticed in these graphs. A finer dicretization might give more information on occuring patterns, but will lead to a long computational time. Figure E.9, and Figure E.10 show the magnitude of the sensitivity indices of two different second-order effects with fine discretization steps of 0.05 for γ and 0.025 for α. Figure E.9 is a finer discretization of Figure E.7, and Figure E.10 is a finer discretization of Figure E.8.

Figure E.9a shows that the average sensitivity index on queue time between γ and α is approximately 0.4. It is seen that for some specific settings it can reach a value of 0.8. An explanation for this was not found. The sensitivity indices on hit rate for the two parameters, seen in Figure E.9b, shows that the secondary effects are smaller. It averages around 0.3, and has identifiable regions for low and high α where the sensitivity indices are below and above average. Figure E.10a, and Figure E.10b, show less variation in values of sensitivity indices, and are on average 0.4 for both the effect on queue time and hit rate. However, some exceptions are also present in these graphs, where an index value above 0.7 is reached. Overall, secondary effects have a considerable impact on the variance of queue time and hit rate.

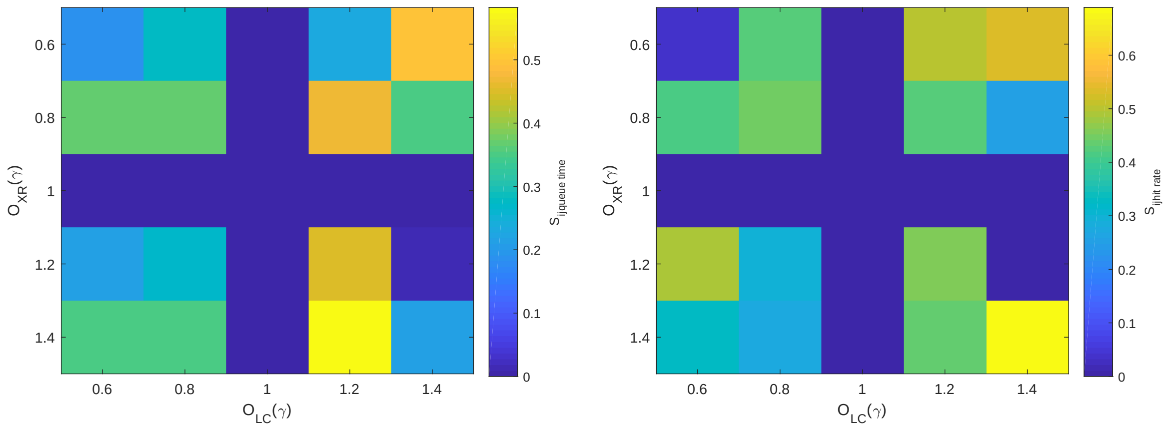
For this experiment, it was expected that certain relations of efficiency and security could be extracted between parameters of operators. However, the sensitivity indices indicate that characterization of any relations between γ and α is difficult. That indicates that some emergent occurring behaviour could be present, which can be difficult to understand, especially without prior expectations.



(a) The sensitivity index for queuing time.

(b) The sensitivity index for hit rate.

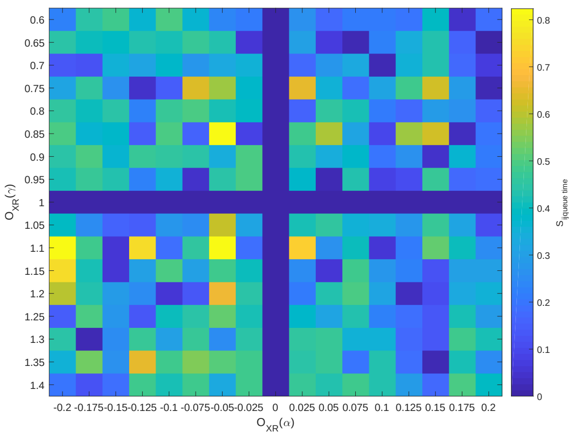
Figure E.7: Sensitivity indices of the second-order effect of γ of the X-Ray operator with respect to the α of the X-Ray operator.



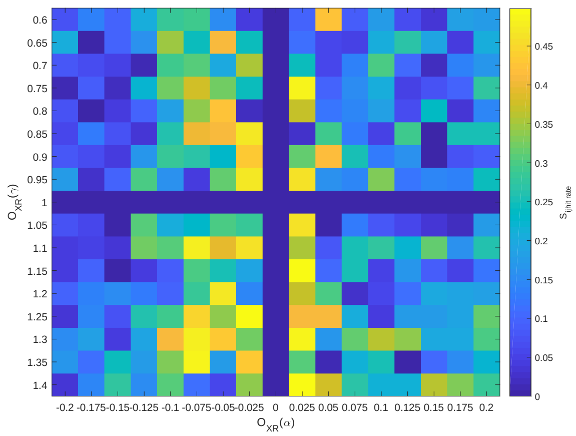
(a) The sensitivity index for queuing time.

(b) The sensitivity index for hit rate.

Figure E.8: Sensitivity indices of the second-order effect of γ of the X-Ray operator with respect to the γ of the luggage check operator.

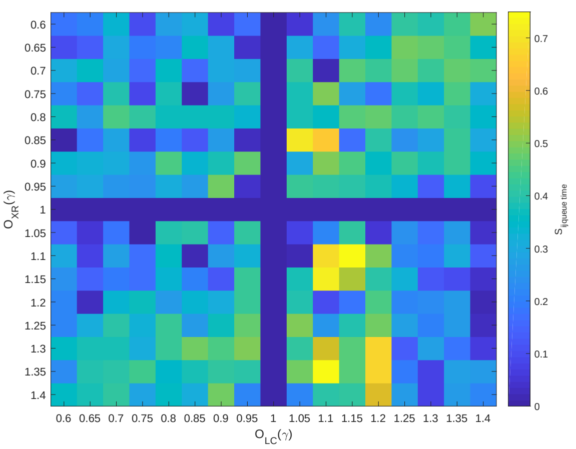


(a) The sensitivity index for queuing time.

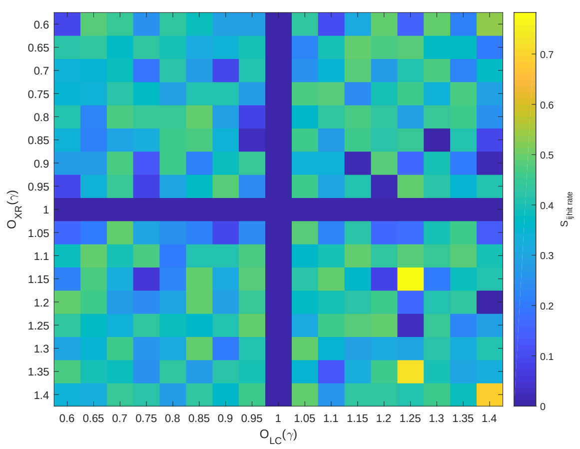


(b) The sensitivity index for hit rate.

Figure E.9: Sensitivity indices of the second-order effect of γ of the X-Ray operator with respect to the α of the X-Ray operator.



(a) The sensitivity index for queuing time.



(b) The sensitivity index for hit rate.

Figure E.10: Sensitivity indices of the second-order effect of γ of the X-Ray operator with respect to the γ of the luggage check operator.

F
SURROGATE

The case study in this research is concerned with choosing the best alternative out of the scheduling alternatives. The case study has a total of 675 scheduling alternatives which exceeds the computational cost if each alternative has to be sampled 500 times per slot. In practice, if a weekly schedule is considered with all of the available operators, the number of alternatives would be very high. Surrogate modelling allows to create a model that approximates the behaviour of the agent-based model while reducing the computational cost. Building a surrogate model of an agent-based model as used in this study for practical purposes is a complex task as the airport terminal is a high-dimensional parameter space where high-order interaction effects occur. Since the surrogate is not the main focus of the research objective, a simple additive surrogate was built, for the specific departure of flights as seen in Table F.1, that is able to replicate the impact of operator accuracy, speed, and fatigue, on the output parameters queue time and hit rate.

Table F.1: Scheduled departing flights used in the agent-based model for the case study. Flights 1 to 6 are in the morning slot, flights 7 to 9 are in the afternoon slot.

Flight	Nr. of passengers	Departure time
f1	70	07:05
f2	130	07:05
f3	150	07:10
f4	140	07:25
f5	130	07:30
f6	135	07:35
f7	140	15:50
f8	140	16:25
f9	150	16:45

F.1. METHOD

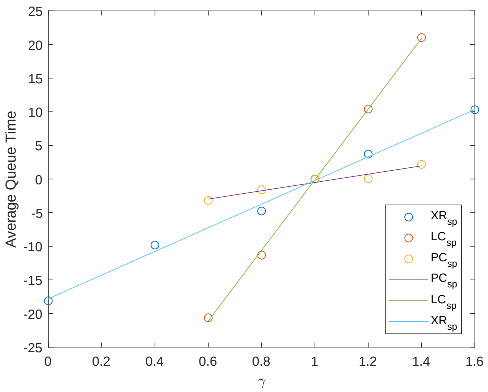
The assumption is made that the addition of first-order effects of operators is a good approximation for the outputs queue time and hit rate. Some tests with the speed parameter γ and accuracy shift parameter α,were performed in which second-order and third-order effects could be approximated within a 5% error margin by adding up first-order effects. However, proving this for all combinations of parameters, for the specified schedule, would require too large computational efforts.

A baseline is set which indicates the average situation for the morning slot and the afternoon slot. The morning slot has 4 open lanes, whereas the afternoon slot has 2. The baseline includes operators defined as $O(\gamma,\alpha,\sigma,\phi) = O(1.0,-0.15,0.95,8)$. This baseline is kept constant while a single parameter of a single operator is changed in one of the slots. Table F.2 shows the discretization scheme of the explored parameters. This result of the simulation outputs is subtracted by the base value for average queue time or hit rate. These points are then interpolated such that all points in between can be selected to specify the operators in the surrogate. The surrogate is then able to return the average queue time and hit rate for each possible combination of defined operators.

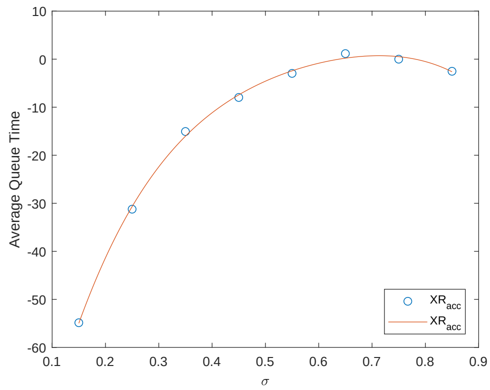
Figure F.1, and Figure F.2 show the first-order effects of changing γ, σ, and ϕ. These figures show the results which are used in the surrogate model. The y-axes show how much the average queue time or hit rate is affected if an operator is specified by the value on the x-axes. The impact of $O_{lc}(\sigma)$ and $O_{lc}(\sigma)$ are assumed negligible. Furthermore, fatigue has a negligible effect on average queue time.

Table F.2: Discretization scheme of simulated parameters.

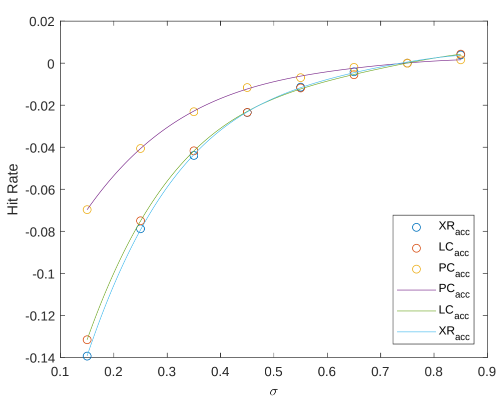
Variable	Discretization	Response time / TPR
$O_{xr}(\gamma)$	[0:0.4:1.6]	[4.50 5.97 7.44 8.90 10.37]
$O_{lc}(\gamma)$	[0.6:0.2:1.4]	[41.58 53.94 66.30 78.66 91.02]
$O_{pc}(\gamma)$	[0.6:0.2:1.4]	[24.12 30.66 37.20 43.74 50.28]
$O_{xr}(\sigma)$	[0.15:0.1:0.85]	[0.29 0.57 0.75 0.82 0.89 0.92 0.95 0.96]
$O_{lc}(\sigma)$	[0.15:0.1:0.85]	[0.29 0.57 0.75 0.82 0.89 0.92 0.95 0.96]
$O_{pc}(\sigma)$	[0.15:0.1:0.85]	[0.29 0.57 0.75 0.82 0.89 0.92 0.95 0.96]
$O_{xr}(\phi)$	[4:2:8]	-
$O_{lc}(\phi)$	[4:2:8]	-
$O_{pc}(\phi)$	[4:2:8]	-



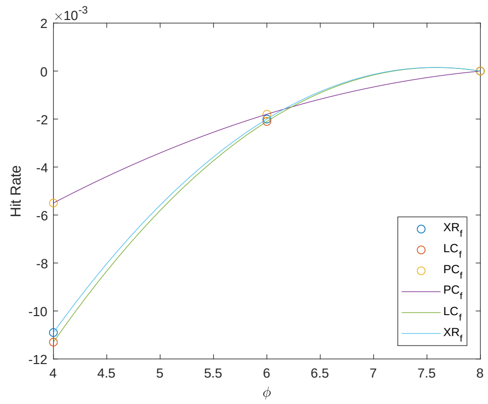
(a) Average queue time impact of speed parameter γ



(b) Average queue time impact of accuracy parameter σ

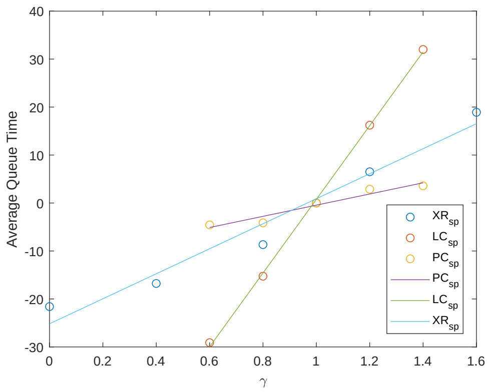


(c) Hit rate impact of accuracy parameter σ

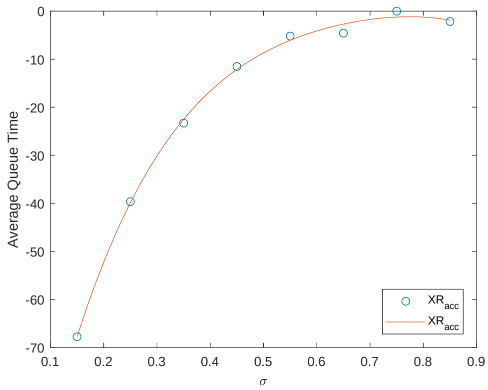


(d) Average queue time impact of fatigue parameter ϕ

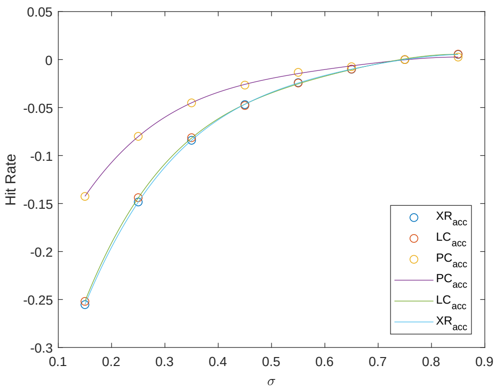
Figure F.1: First-order effect of changing speed, accuracy, or fatigue of an operator on slot 1, the morning slot.



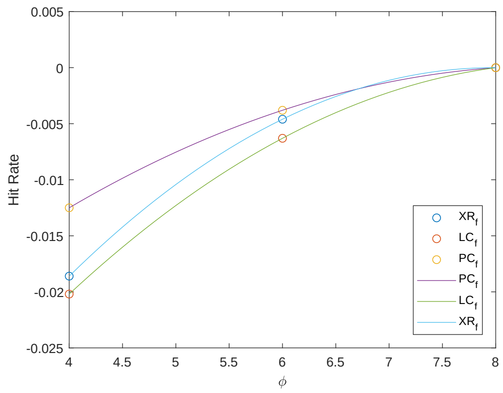
(a) Average queue time impact of speed parameter γ



(b) Average queue time impact of accuracy parameter σ



(c) Hit rate impact of accuracy parameter σ



(d) Average queue time impact of fatigue parameter ϕ

Figure F.2: First-order effect of changing speed, accuracy, or fatigue of an operator on slot 2, the afternoon slot.

Bibliography

[1] A. Washington, *All-hazards risk and resilience: prioritizing critical infrastructures using the ramcap plus [hoch] sm approach,* (ASME, 2009).

[2] J. P. Brashear and J. W. Jones, *Risk analysis and management for critical asset protection (ramcap plus),* Wiley handbook of science and technology for homeland security , 1 (2008).

[3] L. A. Cox, Jr, *Some limitations of "risk= threat× vulnerability× consequence" for risk analysis of terrorist attacks,* Risk Analysis: An International Journal **28**, 1749 (2008).

[4] S. Janssen and A. Sharpanskykh, *Agent-based modelling for security risk assessment,* in *International Conference on Practical Applications of Agents and Multi-Agent Systems* (Springer, 2017) pp. 132–143.

[5] S. Janssen, A. Sharpanskykh, and R. Curran, *Absrim: An agent-based security risk management approach for airport operations,* Risk Analysis (2019).

[6] A. Knol, *Analysing security and efficiency of airport terminal operations, literature review,* (2018).

[7] T. Ulku, *Efficiency of german airports and influencing factors,* Master Diploma of Humboldt University (2009).

[8] T. Hentschel, *Indicators and measures to enhance operational efficiency for regional airports,* ACI Europe's Regional Airports Forum 4th Annual Conference (2011).

[9] K. E. Yoo and Y. C. Choi, *Analytic hierarchy process approach for identifying relative importance of factors to improve passenger security checks at airports,* Journal of Air Transport Management **12**, 135 (2006).

[10] A. A. Kirschenbaum, *The cost of airport security: The passenger dilemma,* Journal of Air Transport Management **30**, 39 (2013).

[11] D. H. Harris, *How to really improve airport security,* Ergonomics in Design **10**, 17 (2002).

[12] A. Schwaninger, *Increasing efficiency in airport security screening,* WIT Transactions on the Built Environment **82** (2005).

[13] A. A. Kirschenbaum, M. Mariani, C. Van Gulijk, S. Lubasz, C. Rapaport, and H. Andriessen, *Airport security: An ethnographic study,* Journal of air transport management **18**, 68 (2012).

[14] A. Schwaninger, *Airport security human factors: From the weakest to the strongest link in airport security screening.* in *Proceedings of the 4th International Aviation Security Technology Symposium* (2006).

[15] İ. Korkmaz, H. Gökçen, and T. Çetinyokuş, *An analytic hierarchy process and two-sided matching based decision support system for military personnel assignment,* Information Sciences **178**, 2915 (2008).

[16] P. P.-Y. Wu and K. Mengersen, *A review of models and model usage scenarios for an airport complex system,* Transportation Research Part A: Policy and Practice **47**, 124 (2013).

[17] S. Eilon and S. C. Mathewson, *A simulation study for the design of an air terminal building,* in *Proceedings of the 5th conference on Winter simulation* (ACM, 1971) pp. 278–294.

[18] L. Brunetta, L. Righi, and G. Andreatta, *An operations research model for the evaluation of an airport terminal: Slam (simple landside aggregate model),* Journal of Air Transport Management **5**, 161 (1999).

[19] M. Campanella, S. Hoogendoorn, and W. Daamen, *The nomad model: theory, developments and applications,* Transportation Research Procedia **2**, 462 (2014).

[20] D. Wilson, E. K. Roe, and S. A. So, *Security checkpoint optimizer (sco): an application for simulating the operations of airport security checkpoints,* in *Proceedings of the 38th conference on Winter simulation* (Winter Simulation Conference, 2006) pp. 529–535.

[21] G. Perboli, S. Musso, F. Perfetti, and P. Trapani, *Simulation of new policies for the baggage check in the security gates of the airports: the logiscan case study,* Procedia-Social and Behavioral Sciences **111**, 58 (2014).

[22] P. K. Chawdhry, *Risk modeling and simulation of airport passenger departures process,* in *Winter Simulation Conference* (Winter Simulation Conference, 2009) pp. 2820–2831.

[23] S.-H. Tsaur, T.-Y. Chang, and C.-H. Yen, *The evaluation of airline service quality by fuzzy mcdm,* Tourism management **23**, 107 (2002).

[24] M. Barbati, G. Bruno, and A. Genovese, *Applications of agent-based models for optimization problems: A literature review,* Expert Systems with Applications **39**, 6020 (2012).

[25] D. K. Huang, H. N. Chiu, R. H. Yeh, and J. H. Chang, *A fuzzy multi-criteria decision making approach for solving a bi-objective personnel assignment problem,* Computers & Industrial Engineering **56**, 1 (2009).

[26] M. Sabar, B. Montreuil, and J.-M. Frayret, *A multi-agent-based approach for personnel scheduling in assembly centers,* Engineering Applications of Artificial Intelligence **22**, 1080 (2009).

[27] M. Sabar, B. Montreuil, and J.-M. Frayret, *An agent-based algorithm for personnel shift-scheduling and rescheduling in flexible assembly lines,* Journal of Intelligent Manufacturing **23**, 2623 (2012).

[28] S. Janssen, *Agent-based security and efficiency estimation in airport terminals,* in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems* (International Foundation for Autonomous Agents and Multiagent Systems, 2017) pp. 1840–1841.

[29] B. Meijer, *Agent-based security risk assessment of the airport security system,* (2017).

[30] A.-N. Blok, *Towards formalization and operationalization of resilience,* (2018).

[31] A. van den Berg, *Quantifying vulnerabilities in an airport checkpoint: A study on the role of employee behavior in the emergence of vulnerabilities,* (2018).

[32] A. Knol, *Modelling and analysis of dynamic relations and trade-offs between security and efficiency in airport security operations,* (2018).

[33] M. Jesus, *Trade-off analysis between security and efficiency of airport operations: A methodological underpinned modelling approach to identify, analyse and evaluate trade-offs based on empirical choice observations,* (2018).

[34] S. Janssen, A.-N. Blok, and A. Knol, *Aatom—an agentbased airport terminal operations model,* (2018).

[35] A. Guitouni and J.-M. Martel, *Tentative guidelines to help choosing an appropriate mcda method,* European Journal of Operational Research **109**, 501 (1998).

[36] M. Velasquez and P. T. Hester, *An analysis of multi-criteria decision making methods,* International Journal of Operations Research **10**, 56 (2013).

[37] D. L. Olson, *Decision aids for selection problems* (Springer Science & Business Media, 1996).

[38] M. B. Barfod and S. Leleur, *Multi-criteria decision analysis for use in transport decision making,* (2014).

[39] M. Marttunen, J. Lienert, and V. Belton, *Structuring problems for multi-criteria decision analysis in practice: A literature review of method combinations,* European Journal of Operational Research **263**, 1 (2017).

[40] A. Görener, *Comparing ahp and anp: An application of strategic decisions making in a manufacturing company,* International Journal of Business and Social Science **3** (2012).

[41] E. Roszkowska, *Multi-criteria decision making models by applying the topsis method to crisp and interval data,* Multiple Criteria Decision Making/University of Economics in Katowice **6**, 200 (2011).

[42] S. Opricovic and G.-H. Tzeng, *Compromise solution by mcdm methods: a comparative analysis of vikor and topsis,* European Journal of Operational Research **156**, 445 (2004).

[43] M. Behzadian, R. B. Kazemzadeh, A. Albadvi, and M. Aghdasi, *Promethee: A comprehensive literature review on methodologies and applications,* European journal of Operational research **200**, 198 (2010).

[44] J.-P. Brans, P. Vincke, and B. Mareschal, *How to select and how to rank projects: The promethee method,* European journal of operational research **24**, 228 (1986).

[45] J. Figueira, V. Mousseau, and B. Roy, *Electre methods,* in *Multiple criteria decision analysis: State of the art surveys* (Springer, 2016) pp. 133–153.

[46] M. G. Yücel and A. Görener, *Decision making for company acquisition by electre method,* International Journal of Supply Chain Management **5**, 75 (2016).

[47] W.-C. Huang and C.-H. Chen, *Using the electre ii method to apply and analyze the differentiation theory,* in *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol. 5 (2005) pp. 2237–2249.

[48] E. Triantaphyllou and S. H. Mann, *An examination of the effectiveness of multi-dimensional decision-making methods: a decision-making paradox,* Decision Support Systems **5**, 303 (1989).

[49] E. Kornyshova and C. Salinesi, *Mcdm techniques selection approaches: state of the art,* in *Computational Intelligence in Multicriteria Decision Making, IEEE Symposium on* (IEEE, 2007) pp. 22–29.

[50] C. Kahraman, *Fuzzy multi-criteria decision making: theory and applications with recent developments*, Vol. 16 (Springer Science & Business Media, 2008).

[51] L. A. Zadeh, *A note on z-numbers,* Information Sciences **181**, 2923 (2011).

[52] L. A. Zadeh, *Stratification, target set reachability and incremental enlargement principle,* Information Sciences **354**, 131 (2016).

[53] M. R. Asadabadi, *The stratified multi-criteria decision-making method,* Knowledge-Based Systems (2018).

[54] J. Rezaei, *Best-worst multi-criteria decision-making method,* Omega **53**, 49 (2015).

[55] D. Pamučar, Ž. Stević, and S. Sremac, *A new model for determining weight coefficients of criteria in mcdm models: Full consistency method (fucom),* Symmetry **10**, 393 (2018).

[56] T. Saaty and D. Ergu, *When is a decision-making method trustworthy? criteria for evaluating multi-criteria decision-making methods,* International Journal of Information Technology Decision Making **46**, 150811195945009 (2015).

[57] R. Karni, P. Sanchez, and V. R. Tummala, *A comparative study of multiattribute decision making methodologies,* Theory and Decision **29**, 203 (1990).

[58] P. Salminen, J. Hokkanen, and R. Lahdelma, *Comparing multicriteria methods in the context of environmental problems,* European Journal of Operational Research **104**, 485 (1998).

[59] P.-L. Lai, A. Potter, M. Beynon, and A. Beresford, *Evaluating the efficiency performance of airports using an integrated ahp/dea-ar technique,* Transport Policy **42**, 75 (2015).

[60] P. M. da Rocha, A. P. de Barros, G. B. da Silva, and H. G. Costa, *Analysis of the operational performance of brazilian airport terminals: A multicriteria approach with de borda-ahp integration,* Journal of air transport management **51**, 19 (2016).

[61] A. Mardani, A. Jusoh, K. MD Nor, Z. Khalifah, N. Zakwan, and A. Valipour, *Multiple criteria decision-making techniques and their applications–a review of the literature from 2000 to 2014,* Economic Research-Ekonomska Istraživanja **28**, 516 (2015).

[62] H. Aboutorab, M. Saberi, M. R. Asadabadi, O. Hussain, and E. Chang, *Zbwm: The z-number extension of best worst method and its application for supplier development,* Expert Systems with Applications **107**, 115 (2018).

[63] A. Knol, A. Sharpanskykh, and S. Janssen, *Analyzing airport security checkpoint performance using cognitive agent models,* Journal of Air Transport Management **75**, 39 (2019).

[64] J. Skorupski and P. Uchroński, *A fuzzy model for evaluating airport security screeners' work,* Journal of Air Transport Management **48**, 42 (2015).

[65] F. Hofer and A. Schwaninger, *Using threat image projection data for assessing individual screener performance,* WIT Transactions on the Built Environment **82** (2005).

[66] J. S. McCarley, A. F. Kramer, C. D. Wickens, E. D. Vidoni, and W. R. Boot, *Visual skills in airport-security screening,* Psychological science **15**, 302 (2004).

[67] R. Ratcliff and J. N. Rouder, *Modeling response times for two-choice decisions,* Psychological science **9**, 347 (1998).

[68] T. V. Wiecki, I. Sofer, and M. J. Frank, *Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python,* Frontiers in neuroinformatics **7**, 14 (2013).

[69] P. McCauley, L. V. Kalachev, D. J. Mollicone, S. Banks, D. F. Dinges, and H. P. Van Dongen, *Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance,* Sleep **36**, 1987 (2013).

[70] A. Schwaninger, *Reliable measurements of threat detection.* (2003).

[71] J. Vandekerckhove and F. Tuerlinckx, *Diffusion model analysis with matlab: A dmat primer,* Behavior research methods **40**, 61 (2008).

[72] M. M. Walsh, G. Gunzelmann, and H. P. Van Dongen, *Computational cognitive modeling of the temporal dynamics of fatigue from sleep loss,* Psychonomic bulletin & review **24**, 1785 (2017).

[73] K. Leone and R. R. Liu, *Improving airport security screening checkpoint operations in the us via paced system design,* Journal of Air Transport Management **17**, 62 (2011).

[74] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, *Normalization techniques for multi-criteria decision making: analytical hierarchy process case study,* in *doctoral conference on computing, electrical and industrial systems* (Springer, 2016) pp. 261–269.

[75] S. Howard, I. A. Scott, H. Ju, L. McQueen, and P. A. Scuffham, *Multicriteria decision analysis (mcda) for health technology assessment: the queensland health experience,* Australian Health Review (2018).

[76] K. Marsh, M. IJzerman, P. Thokala, R. Baltussen, M. Boysen, Z. Kaló, T. Lönngren, F. Mussen, S. Peacock, J. Watkins, *et al.*, *Multiple criteria decision analysis for health care decision making—emerging good practices: report 2 of the ispor mcda emerging good practices task force,* Value in health **19**, 125 (2016).

[77] R. E. Nisbett and T. D. Wilson, *Telling more than we can know: Verbal reports on mental processes.* Psychological review **84**, 231 (1977).

[78] L. J. Mangham, K. Hanson, and B. McPake, *How to do (or not to do)... designing a discrete choice experiment for application in a low-income country,* Health policy and planning **24**, 151 (2009).

[79] M. Ben-Akiva and M. Bierlaire, *Discrete choice methods and their applications to short term travel decisions,* in *Handbook of transportation science* (Springer, 1999) pp. 5–33.

[80] C. G. Chorus, T. A. Arentze, and H. J. Timmermans, *A random regret-minimization model of travel choice,* Transportation Research Part B: Methodological **42**, 1 (2008).

[81] R. M. Johnson and B. K. Orme, *How many questions should you ask in choice-based conjoint studies,* in *Art Forum, Beaver Creek* (1996) pp. 1–23.

[82] E. W. de Bekker-Grob, B. Donkers, M. F. Jonker, and E. A. Stolk, *Sample size requirements for discrete-choice experiments in healthcare: a practical guide,* The Patient-Patient-Centered Outcomes Research **8**, 373 (2015).

[83] K. E. Train, *Discrete choice methods with simulation* (Cambridge university press, 2009).

[84] J. M. Rose and M. C. Bliemer, *Constructing efficient stated choice experimental designs,* Transport Reviews **29**, 587 (2009).

[85] S. van Cranenburgh and A. T. Collins, *New software tools for creating stated choice experimental designs efficient for regret minimisation and utility maximisation decision rules,* Journal of Choice Modelling **31**, 104 (2019).

[86] J. J. Louviere, D. A. Hensher, and J. D. Swait, *Stated choice methods: analysis and applications* (Cambridge university press, 2000).

[87] B. Orme, *Interpreting the results of conjoint analysis,* Getting Started with Conjoint Analyis: Strategies for Product Design and Pricing Research **2**, 77 (2010).

[88] F. E. Ritter, M. J. Schoelles, K. S. Quigley, and L. C. Klein, *Determining the number of simulation runs: Treating simulations as theories by not sampling their behavior,* in *Human-in-the-loop simulations* (Springer, 2011) pp. 97–116.

[89] I. Lorscheid, B.-O. Heine, and M. Meyer, *Opening the 'black box'of simulations: increased transparency and effective communication through the systematic design of experiments,* Computational and Mathematical Organization Theory **18**, 22 (2012).

[90] P. M. Krueger and E. M. Friedman, *Sleep duration in the united states: a cross-sectional population-based study,* American journal of epidemiology **169**, 1052 (2009).

[91] B. Iooss and P. Lemaître, *A review on global sensitivity analysis methods,* in *Uncertainty management in simulation-optimization of complex systems* (Springer, 2015) pp. 101–122.

[92] J.-S. Lee, T. Filatova, A. Ligmann-Zielinska, B. Hassani-Mahmooei, F. Stonedahl, I. Lorscheid, A. Voinov, J. G. Polhill, Z. Sun, and D. C. Parker, *The complexities of agent-based modeling output analysis,* Journal of Artificial Societies and Social Simulation **18**, 4 (2015).

[93] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity analysis in practice: a guide to assessing scientific models,* Chichester, England (2004).