

Probabilistic Runway and Capacity Forecasting using Machine Learning to Support Decision Making

Arjen de Leege
To70
The Hague, The Netherlands
arjen.deleege@to70.nl

Ceriel Janssen
KLM Royal Dutch Airlines
Amstelveen, The Netherlands
ceriel.janssen@klm.com

In this paper we present a system that supports airlines to take timely actions to mitigate the impact of runway capacity shortfalls and unforeseen long taxi times on airline performance, and to optimize fuel on board for airborne holding and taxiing. The system makes use of a machine learning technique to provide a 30-hour probabilistic forecast. The system has been validated to ascertain its predictive power and to determine the impact on decision making. The probabilistic forecast matches with the realized use fractions to within 7%. An experiment in using the display indicates that decisions to cancel flights are made earlier which allows more time to re-route passengers, and better targeting of advices to take extra fuel on board to allow for airborne holding.

Decision Making, Decision Support, Machine Learning, Runway Use, Runway Capacity, Forecasting,

I. INTRODUCTION

Airlines suffer financial and reputational damage when passengers miss their connecting flight due to airline non-performance [1]. Therefore, timely actions are key for airlines when airline performance is at risk.

Runway capacity shortfalls and selection of runways that lead to unforeseen long taxi times have a detrimental effect on airline performance. Weather is a determining factor in runway selection and runway capacity. However, uncertainty of the weather forecast complicates forecasting of the runways that will be used and the capacity which they provide. Additionally, runway selection varies from day to day under influence of factors that cannot easily be captured or modelled for inclusion in a runway forecast.

Amsterdam Airport Schiphol (AAS) has six runways that are used in more than eighty different combinations. During off-peak hours two runways are used: one for take-offs and one for landings. During inbound and outbound peaks three runways are used. The third runway is used for either take-offs or landings. During the day, Schiphol alternates between inbound and outbound peaks, resulting in a minimum of 14 runway configuration changes per day. Between an inbound or outbound peak, four runways may be used simultaneously for a

short period of time (double peak). Finally, runways are used in a noise preferential order [2].

The variety of runway combinations, high amount of changes, and complexity of the runway selection criteria limit the ability of airlines to accurately predict airport capacity and taxi times. Various runway use forecasting systems have therefore been developed [3,4,5]. In this paper a novel approach is taken by application of machine learning to derive a predictive model and method of presentation of the results to support decision making.

This paper presents a system that provides a probabilistic forecast of runway use and capacity for the next 30-hours in 20 minute intervals. The system uses a machine-learning technique to forecast runway use based on weather observations and prior observed runway selection. Monte Carlo simulations are applied with varying weather conditions that are derived from a probabilistic meteorological forecast. This approach provides a probabilistic forecast of runway use and capacity that accounts for both the uncertainty in the weather and uncertainty in runway selection. The forecast is made available to end users using an online dashboard.

The remainder of this paper is organized as follows: Section II introduces the machine learning approach to runway and capacity forecasting. Section III discuss the process to generate the forecast. Section IV discusses how the probability forecast is presented to the end users. In Section V the results of the analysis to determine the predictive power of the system are presented. In Section VI discussed the impact on the decision making process. Sections VII and VIII contain the discussion and conclusion.

II. MACHINE LEARNING APPROACH TO RUNWAY FORECASTING

A. Machine Learning

Runway selection at AAS is based on required arrival capacity, required departure capacity, crosswind and tailwind limits, minimum cloud base, runway visual range, and noise preferential runway use. Explicit modelling the runway

This research was sponsored by the Knowledge Development Centre Mainport Schiphol.

combination selection process based on these inputs would lead to a highly complex model.

Machine learning is defined by Mitchell in Ref. [6] as follows: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .”. Using machine learning techniques, a runway forecast can be made without explicit modeling of the runway selection.

Meteorological observations (Meteorological Aerodrome Reports, METAR) and runway use data are used to train a predictive model. During the training phase model parameters are set by a learning algorithm. Training ends when the model parameter changes drop below a preset threshold value. After the model is trained, the model is ready for use. Figure 1 gives an overview of the proposed approach.

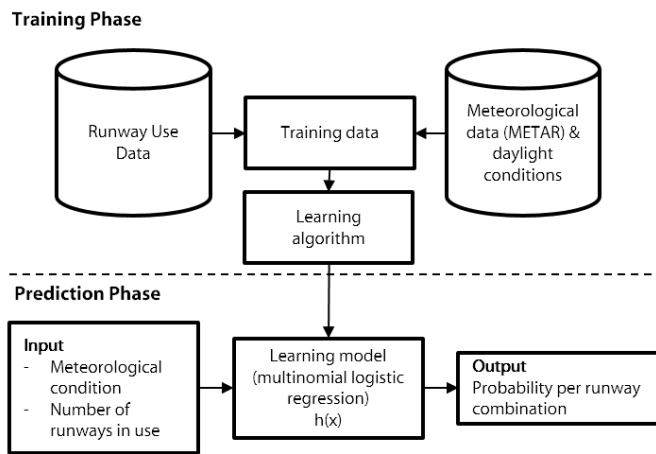


Figure 1. Schematic overview of training and prediction processes.

B. Multinomial Logistic Regression

In our machine learning approach, we define runway forecasting as a supervised learning regression problem. We use multinomial logistic regression to predict the probabilities of all possible runway combinations that can be used. Multinomial logistic regression is well described and readily available in tools and libraries like Hadoop and Weka [7,8,9]. Therefore, our discussion is limited to its application in forecasting runway use.

The runway combination is considered a nominal dependent variable. The number of categories is limited (i.e. 82 unique runway combinations), there is no ordering in any meaningful order and all categories are known.

C. Model Features

To predict runway use, the model uses the features (or inputs) listed in Table I. The features include the determining

factors for runway use: wind, visibility, the daylight condition, and the number of runways in use.

Table 1. Model Features

Features	Unit / Categories
Visibility condition	Good / Marginal / Low Visibility Procedure phase A to C
Uniform Daylight Period	Yes / No
Wind speed	kts
Wind direction	kts
Gust	kts
Variation wind direction >60°	Yes / No
Number of runways in use	1+1, 1+2,2+1,2+2*

*Landing + Take-Off

D. Runway Capacity

The capacity of individual runways depends on the separation criteria, actual fleet mix, and headwind. ATC declares an hourly capacity per runway combination that has been based on these parameters. The capacity is given per runway as function of the combination in use, visibility, and daylight conditions. To determine the runway capacity, the model derives the capacity from the forecasted runway combination using this declared capacity.

III. PROBABILISTIC RUNWAY AND CAPACITY FORECAST

As discussed in Section II the runway use model predicts the probabilities of all possible runway combinations based on the expected meteorological condition and the number of runways in use. To generate a probabilistic forecast that also includes the uncertainty in the weather, Monte Carlo simulations are performed with varying meteorological conditions. The meteorological conditions are derived from the Schiphol Probability Forecast (SKV, Dutch: Schiphol Kansverwachting).

KNMI PROBABILITY FORECAST SCHIPHOL
Friday 21 November 12 UTC till Saturday 22 November 18 UTC
Last update: Short term: 09:45 UTC Long term: 10:57 UTC

	12	13	14	15	16	17	18	21	00	03	06	09	12	15	18
Visibility < 5 km and/or ceiling < 1000 ft (%)	0	0	0	0	0	0	0	5	15	20	30	30	30	10	10
RVR < 1500 m and/or ceiling ≤ 300 ft (%)	0	0	0	0	0	0	0	0	0	0	5	5	0	0	5
RVR < 550 m and/or ceiling < 200 ft (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RVR < 350 m (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Winddirection (deg)	090	100	100	100	100	100	090	100	110	120	130	140	140	150	150
Windspeed (kt)	9	10	10	10	10	9	8	9	9	9	10	10	11	10	9
Gusts (kt)	15	15	15	15	15						15	15	16	15	
Standarddeviation winddirection (deg)	15	15	15	15	15	15	15	15	15	15	15	15	15	20	20
Standarddeviation windspeed (kt)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
CB (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Thunderstorm (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Temperature (C)	7	7	8	7	7	6	5	5	5	5	4	6	8	9	9
Dewpoint (C)	3	4	4	3	3	3	3	3	3	3	4	6	7	8	8
Relative humidity (%)	76	81	76	76	76	81	87	87	87	87	100	100	93	93	93
Windchill	4	4	5	4	4	3	2	2	2	2	-0	3	5	6	6
Snow (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Moderate or heavy snow (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Freezing precipitation (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2. Example of Schiphol Probability Forecast, Source: Ref [10].

A. Schiphol Probability Forecast

The SKV is published by the Royal Netherlands Meteorological Institute (KNMI) every hour [10]. The SKV has a prediction horizon of up to 30 hours and a temporal resolution of 1 hour. Figure 2 shows an example of the SKV. The SKV is also available in a machine readable format. The SKV includes, among others, the probability of reduced visibility conditions, the wind direction and wind speed, gust, and standard deviations of wind direction and speed.

B. Monte Carlo Simulations

Figure 3 gives a schematic overview of the Monte Carlo simulations that are executed using the model to generate the forecast. The forecast is generated per 20-minute period starting on the hour. 20-minute periods are used to align with the planning of off-peaks, inbound peaks, and outbound peaks. Simulations start each time a new SKV is received.

At the start of every run, a meteorological condition is drawn randomly from the SKV. A probability forecast is determined for off-, inbound, outbound, and double peak runway use. The forecasts from all runs are aggregated to the runway probability forecast that thus accounts for both the uncertainty in runway selection and the uncertainty in weather.

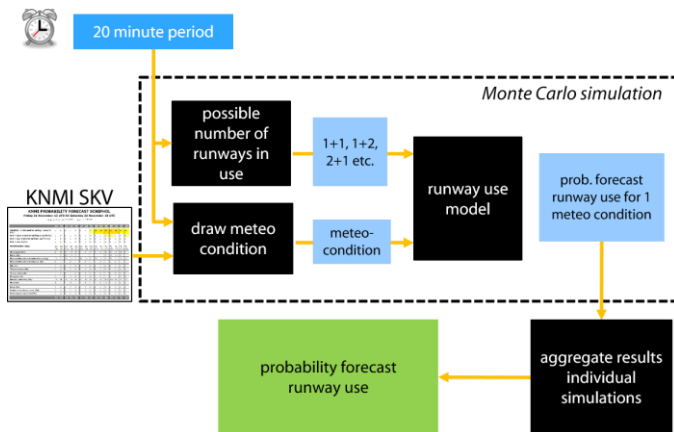


Figure 3. Monte Carlo Simulations run per 20 minute interval.

IV. USER INTERFACE

The probability forecast is published online on a dashboard. The interface has been designed in close operation with its end users at KLM's OCC. Figure 4 gives an impression of the interface.

The system shows an overview of the expected capacity and traffic in the next 30 hours at a glance, but also enables the user to zoom in on the expected runway use, capacity, meteorological conditions, and traffic within a 20-minute period. Furthermore, the interface gives the user the possibility to quickly assess what-if scenarios based on hypothetical conditions.

The interface consists of several blocks that are discussed in the remainder of this section.

A. Capacity – Demand Graph

The graph at the top of the interface shows the capacity per 20-minute for departures (blue) and arrivals (orange) with a probability of 50% or more. The grey bars show the scheduled number of departures and arrivals on the runway. Both capacity and traffic per 20-minute are visualized as movements per hour. Traffic exceeding predicted capacity by a large margin is an indication that runway capacity may be insufficient to handle traffic without delays.

During the day, both the traffic and runway capacity alternate between inbound and outbound peaks. For the runway capacity, the system uses the peak periods set by Air Traffic Control (ATC) in their seasonal planning. If, during an inbound or outbound peak, a runway combination with two runways for arrivals or departure is not feasible, the system falls back to a runway combination with one runway for arrivals and one for departures.

The graph works like a map. By clicking on the graph, the user selects a 20-minute period and the information highlighted or shown in the other blocks changes accordingly.

B. Capacity Table

The block on the left below the capacity vs. demand graph gives a summary of the meteorological condition, the peak period as planned by ATC in their seasonal planning, and the capacity per peak period. The user can select a line in the table after which the same 20-minute period is highlighted in the capacity graph and other blocks are updated accordingly to show information for the selected 20-minute period.

C. Runway Combination Forecast

The block on the right, below the graph shows the runway combination forecast for the 20-period selected by the user. The forecast is given for the peak period set by ATC and the most likely visibility condition. The runway combinations are sort by probability in descending order. Runway combinations with a probability lower than 5% are not shown to the user. The capacity is given per runway in movements per hour.

The user can select another peak period type or visibility condition to see the which runway combinations might be used in those conditions. This selection does not affect any of the other blocks.

D. Meteo Forecast

The block in the bottom right corner shows the meteorological forecast valid for the 20-minute period selected in the capacity vs. demand graph. The information is derived from the SKV that was used to generate the forecast.

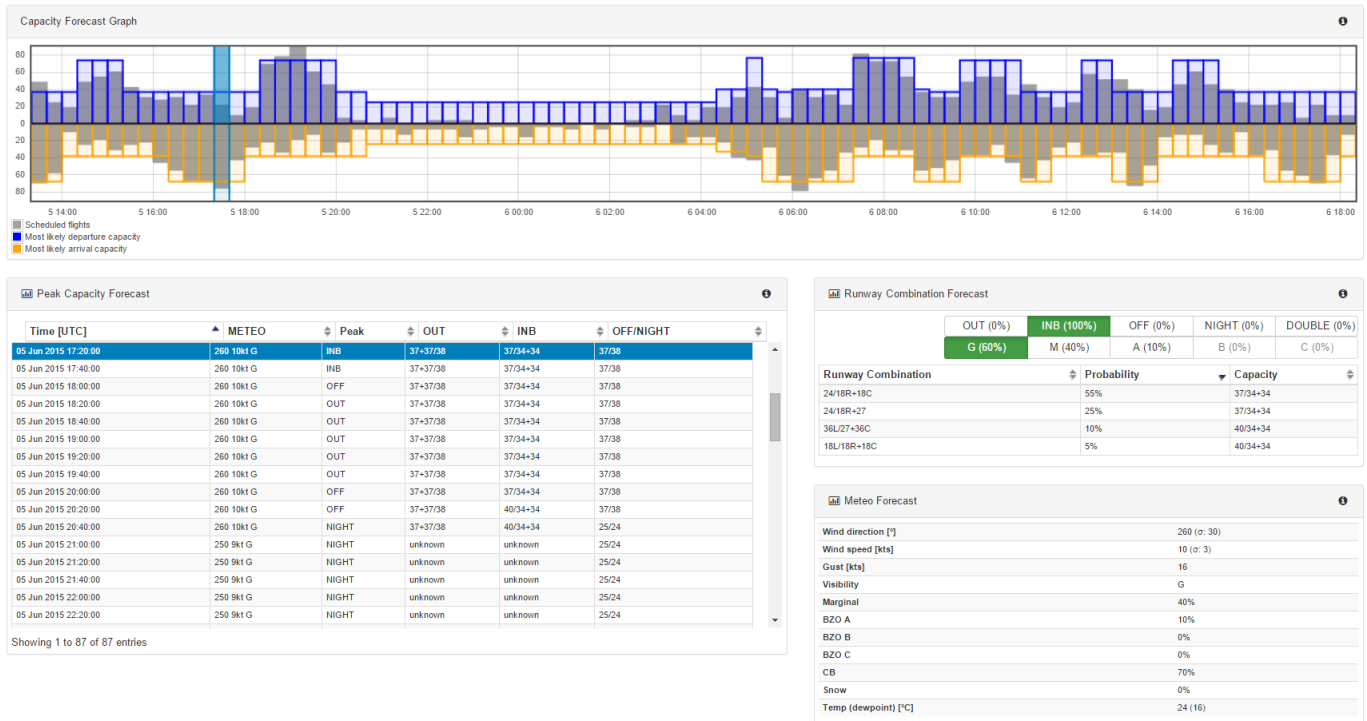


Figure 4. User Interface.

V. PREDICTIVE POWER

The predictive power of the system has been validated by comparing the forecasts to actual runway use. For this purpose, the model was trained using 2013 runway use and METAR data. Subsequently, forecasts were made using 2014 data and compared to the actual realization.

The predictive power of the SKV and runway use model together determine the overall predictive power of the system. The predictive power of the SKV is known [10]. First, the predictive power of the runway use model was determined. Meteorological observations (METAR) were used to forecast the runway use. Second, the overall predictive power of the system was determined. The 2014 archive of the SKV was used to forecast runway use.

A. Predictive Power of Runway Use Model

Figure 5 shows the probability and percentage of time a runway combination was actually used.

The graph shows a strong correlation between probability that a runway combination is used and the percentage that a runway combination is actually used. For example, under those conditions the model forecasts 80% chance a runway combination will be used, the runway combination is used 78% of the time.

Table 2 gives the percentage of time that a runway combination was actually used when the model forecast a 70% chance the runway combination would be used.

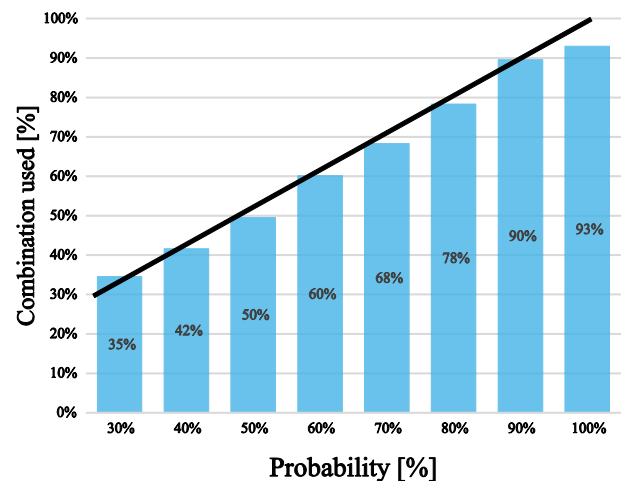


Figure 5. Probability vs. Percentage used.

Table 2. Percentage of time runway combination with a 70% probability actually used per peak period and visibility condition.

Peak period	Good Visibility	Marginal Visibility
Off-peak	72%	64%
Inbound Peak	64%	56%
Outbound Peak	67%	87%
Double Peak	70%	NA
Night	70%	72%

B. Highest Probability in Forecast

The highest probability per forecast is a measure for the predictive power and the added value of the system in the decision-making process. For example, if the maximum probability is 30%, it is likely that the forecast contains multiple runway combinations. This would complicate decision making and limit the added value of the system.

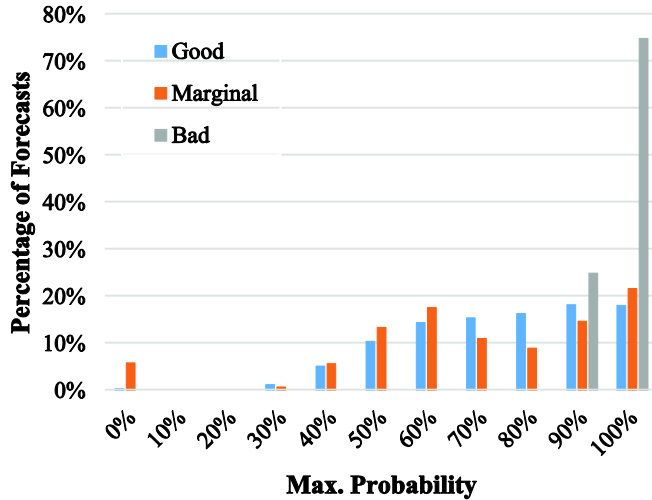


Figure 6. Highest probability per forecast

Figure 6 shows the distribution of the highest probability per forecast. In good visibility, the highest probability is 80% or more in 52% of the forecasts. In general, the highest probability is 40% or less in just 10% of the forecasts.

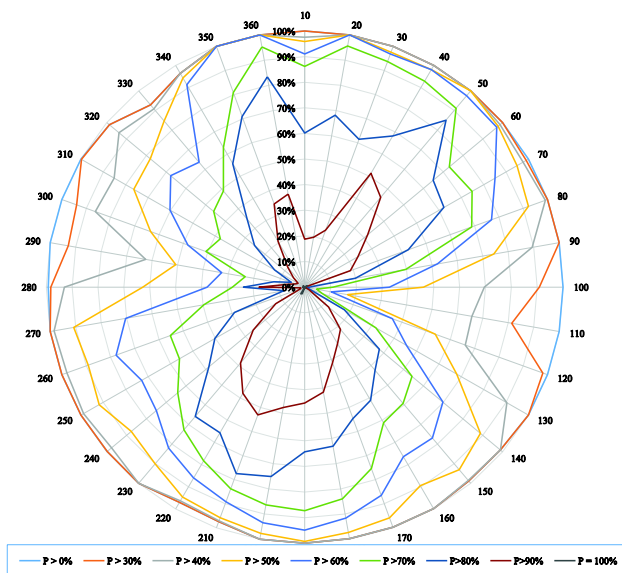


Figure 7. Distribution of highest probability in a forecast vs. wind direction

The highest probability in a forecast strongly depends on the wind direction as shown in Figure 7. If the wind is coming from the East or West, the highest probability in the forecast is significantly lower than for winds from the North or South. This may be explained by the fact that Schiphol has 3 North-South runways which leads to a sensitivity for variation of the headwind components due to winds from the East or West.

C. Predictive Power of the Forecast

Figure 8 shows the percentage of forecasts the runway combination with highest probability (position 1) was used, the percentage the runway combination used was the runway combination with highest or second highest probability (positions 1 & 2) and the percentage of forecasts the runway combination was included in the forecast (probability > 5%).

In 76% of the forecasts made one hour in advance the runway combination used had the highest probability in the forecast. This percentage gradually drops to 69% for forecasts made 27 hours in advance. In 90% of the forecasts made up to 13 hours in advance, the runway combination used had the highest or second highest probability. 90% or more of the forecasts included the runway combination that forecasted with a probability of more than 5%.

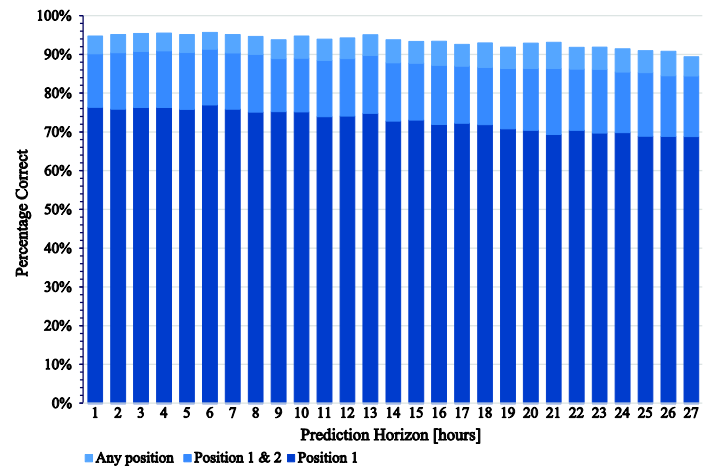


Figure 8. Position of runway combination used in forecast

VI. IMPACT ON DECISION MAKING

To assess the impact of the system on the decision making a comparison was made between the decisions made with and without the system based on actual disturbances.

A. Method

KLM selected five days with (potential) adverse weather conditions that had a negative impact on the performance of the airline. On these days (scenarios), fuel advices were issued and/or flights were canceled.

In an experiment the decision-making process for each of the five scenarios was simulated. The scenarios started at 7:00 the day before and ended at 7:00 on the day itself. Similar to

the actual operation, there are four decision points: 7:00 the day before (D-1), 15:00 D-1, 23:00 D-1, and 7:00 on the day itself. At every decision point, the participant was asked if he or she would like to take any action to mitigate the impact of the weather on the flight operation, based on the forecast made by the system and information from the sector briefing.

The decisions taken during the experiment runs have been compared to the actual decisions that were taken. At the end of the experiment the participants were asked to give feedback on the system and the experiment.

B. Participants

One flow controller and two flight dispatch supervisors participated in the experiment. The participants had no prior experience with the system. For operational reasons and time constraints it was not feasible to invite more participants.

C. Training and Instructions

Prior to the experiment, the participants received a written experiment description. At the start of the experiment an introduction to the system was given using a presentation. Subsequently, the experiment procedure and the roles of the participant and experiment leader were explained. At every decision point the experiment leader provided the latest meteorological forecast and information from the latest sector briefing.

The participants were asked to do the following at every decision point:

1. Observe the information provided and describe the situation;
2. Explain the conclusions from the information provided;
3. Formulate any decisions or actions they would take;
4. Indicate if they would like to move on to the next decision point (8 hours) or reassess the situation within a couple of hours.

One training scenario was used to get acquainted with the system and experiment procedure.

D. Scenarios

The following five days were selected:

Table 3. Scenario Descriptions

Day	Description
1	Southwesterly storm passing over
2	Stormy with winds from an unfavorable direction
3	Snowfall expected, eventually no snowfall on the airport
4	Cold front passing over, combined with strong winds
5	Cold front passing over, combined with very strong winds, gusts up to 50 kts

It took between three and four hours per participant to complete the experiment. The days were presented in a different order to compensate for learning and boredom effects.

E. Impact on Decision Making

In Table 4 a comparison is made between actual decision made on day 4 and the decision made during the experiment. On day 4, a fuel advice had been issued for European flights at 23:00 on the previous day. On the day itself flights had been canceled. Two of the three participants, with the support of the system, took the decision to cancel flights on the day before in the afternoon followed by fuel advices. All participants issued a fuel advice to the European (EUR) flights, and two also issued such advice to intercontinental (ICA) flights. One participant only issued fuel advices. A similar comparison has been made for the 4 four other days. Most of the time the 'type' of decision made in the experiment with the use of the system matched the decision actually taken without the system.

Table 4. Decisions compared for scenario 4.

Decision Point	Actual	Participant		
		1	2	3
7:00 D-1	NONE	NONE	NONE	NONE
15:00 D-1	NONE	FUEL - ICA	CANCEL	CANCEL
23:00 D-1	FUEL - EUR	NONE	FUEL - EUR	FUEL - ICA
7:00	CANCEL	FUEL - EUR	NONE	FUEL - EUR

However, the comparison indicates that the system does affect the moment the decision is taken. Tables 5 and 6 give an overview of the number of participant who decided to cancel flights or issue a fuel advice and how many times a decision was made at an earlier, the same, or later decision point.

Table 5. Decisions to cancel flights compared

Day	Actual	Experiment	Earlier	Same	Later
1	Yes	1/3	0	0	1
2	Yes	3/3	0	2	1
3	Yes	3/3	3	0	0
4	Yes	2/3	2	0	0
5	No	2/3	NA	NA	NA
Sum			5	2	2

Table 6. Decisions to issue fuel advices compared

Day	Actual	Experiment	Earlier	Same	Later
1	Yes	3/3	2	1	0
2	Yes	3/3	0	0	3
3	Yes	3/3	0	0	3
4	Yes	3/3	1	2	0
5	Yes	2/3	0	0	2
Sum			3	3	8

From Table 5 it follows that 5 decisions to cancel flights were taken at an earlier decisions point and two at the same or a later decision point. Contrary to cancelations, Table 6 shows that 8 of the 14 fuel advices were issued later and 3 were issued earlier or at the same time.

The data also suggests that, with the system, the fuel advices are more targeted to short haul and long haul flights arriving within a specific time frame. One of the participants explicitly mentioned this during the experiment.

F. User Feedback and Observations

The participants provided feedback during the experiment and were also asked to provide feedback at the end of the experiment using a questionnaire.

The participants indicated that a capacity shortfall in the top graph works like a trigger to look at in more detail at those periods of the day. Nowadays, they focus primarily on the first inbound peak and last outbound peak of the day. Two participants expect that the system will support them in paying more attention to other peak periods in the middle of the day.

A recurring comment made by the participants was the effort required to get insight in expected weather changes (e.g., backing and veering winds). This requires too many mouse clicks. Two out of three participants missed information about the cross and tailwind per runway and the impact of wind on capacity.

One of the observations made during the experiment was the limited use of capacity table. This was confirmed by the participants and a log of the number of mouse clicks in the table. The participants indicated that the table could be removed from the interface as the other blocks in the interface already contain the information given in the table but in a manner that was easier to understand.

VII. DISCUSSION

In this research we developed a system that provides a probabilistic runway capacity and runway use forecast. The predictive power of the model was validated extensively using two years of data. The probabilities and frequencies deviate no more than 7 percentage points. The runway combination with the highest probability in forecasts made 27 hours in advance

was used 69% of the time. The runway combination with highest or second highest probability in forecasts made 27 hours in advance was used 85% of the time.

While these forecasts do not exactly predict the runway configuration, the forecasts can support decision makers in adjusting their operation. The probabilistic information may not only support changes to operation based on a predicted disruption but also applying appropriate buffers for the likelihood that the disruption may occur.

Unfortunately, the number of people who could participate in the experiment to determine the impact on decision making was limited. At this moment, no statistically significant conclusions can be drawn. Still, the results give a good indication of the impact of the tool on the decision making process. The system primarily affects the moment decisions are taken in a desired way. Decisions to cancel flights are taken earlier, leaving more time to inform and re-route passengers. Decisions to issues fuel advices are taken later and more targeted to short haul or long haul flights arriving in a specific time frame. Hence, overall less extra fuel is taken on board.

The system enables more informed decision making with respect to runway use and runway capacity that can have a detrimental effect on airline performance. However, the information the system provides may also be beneficial to other stakeholders such as the airport, ATC, ground handlers or even people living in the vicinity of the airport. One example is runway maintenance planning. The forecast can also be used to compute the probability a runway will be used the next day. If the probability is low, maintenance is unlikely to interfere with the operation. The runway combination is of less importance. User specific dashboards may be added to the system to cater for the information needs of every stakeholder.

The approach was developed and tested at Amsterdam Airport Schiphol, but may also be applied at other airports. Especially, airlines operating at an airport with a multiple runway system may benefit from the system. To take full benefit of the system a probabilistic weather forecast like the SKV is required.

VIII. CONCLUSIONS

A decision support system was developed to mitigate the impact of runway capacity shortfalls on airline performance by providing a 30-hour probabilistic forecast of runway use and runway capacity. Machine learning was used to derive a predictive model to generate probabilistic runway use forecasts. Combining the probabilistic runway use forecast with a probabilistic meteorological forecast results in a probabilistic forecast that accounts for both the uncertainty in the weather forecast and runway selection. Probabilities forecast have strong correlation with the percentage of time a runway is actually used.

Results indicate that decisions to cancel flights are taken earlier, leaving more time to inform and re-route passengers. Decisions to issues fuel advices are taken later and are more targeted to short haul or long haul flights arriving within a specific timeframe.

ACKNOWLEDGMENT

The authors would like to thank the Knowledge Development Centre (KDC) and KLM for their support. The authors are in particular grateful to the KLM flow controllers and flight dispatchers involved in the design and testing of the system.

REFERENCES

- [1] Mijatovic, D., et al. "Research of the Relation between the Hourly Inbound Capacity at Schiphol Airport and the Number of KLM Transfer Passengers at Risk of Loosing their Connection.", 6th Eurocontrol Innovative Research Workshop & Exhibition, 2007.
- [2] Anon., Integrated Aeronautical Information Package, Air Traffic Control the Netherlands, www.ais-netherlands.nl, retrieved September 2016
- [3] Hesselink, H.H., Nibourg, J.M., "Probabilistic 2-Day Forecast of Runway Use", National Aerospace Laboratory. NLR-TP-2011-287. 2011.
- [4] Ramanujam, V., and Balakrishnan, H., "Estimation of maximum-likelihood discrete-choice models of the runway configuration selection process," in American Control Conference, 2011.
- [5] Houston, S., Murphy, D., "Predicting runway configurations at airports," in Transportation Research Board (TRB) Annual Meeting, 2012, paper number 12-3682.
- [6] Mitchell, T., *Machine Learning*, McGraw-Hill, 1997.
- [7] Greene, W. H., *Econometric Analysis*, 7th Edition, Boston: Pearson Education. pp. 803–806. 2012
- [8] Hall M., Eibe F., Holmes G., Pfahringer B., Reutemann P., Witten I.H., *The WEKA Data Mining Software: An Update*; SIGKDD Explorations, Volume 11, Issue 1. 2009
- [9] Prajapati V., *Big Data Analytics with R and Hadoop*, PACKT, 2013
- [10] KDC - LVP Project team, *Improved Low visibility and Ceiling Forecasts at Schiphol Airport*, Final report, part 1, July 2008.