Using the entropy of multi-aircraft departure time distributions to predict temporal deviations from on-time departure performance

Christoph Moehlenbrink[1] and Alex Kirlik[2]

[1]*German Aerospace Center (DLR), Institute of Flight Guidance*
[2]*University of Illinois Urbana-Champaign, Department of Computer Science*

In airport departure operations, some aircraft departure schedules are more challenging to achieve than others requiring identical numbers of aircraft to depart in a similar amount of time. In particular, some schedules are "stacked," in that they require multiple departures within relatively short time intervals, in contrast to "uniform" schedules in which departure times are relatively evenly spaced throughout the departure window. Metrics capturing the degree to which departure distributions are uniform versus stacked (statistically, uniform versus non-uniform) may provide a quantitative measure useful for predicting on-time departure performance, or equivalently, for predicting departure time deviations. This paper describes how Shannon's entropy metric was used for this purpose. The predictive validity of the measure was evaluated in a medium-fidelity ground control simulation in which human controllers in various experimental conditions were instructed to achieve a timely aircraft departure schedule among other task goals. Regression modeling showed that the entropy of a departure time schedule is a good predictor of the degree to which on-time departure performance was achieved. Correlations between the entropy of departure time distributions and deviations from on-time departure performance of *0*.75 were found. The strengths and limitations of the entropy-based approach to providing a quantitative measure of predicted deviations from on-time departure performance as a function of departure time schedules are discussed.

Keywords – entropy, air traffic control, ground control, human-automation interaction, controller performance

I. INTRODUCTION

Due to the continuous increase in air traffic, there is a need for a better understanding of what human-automation systems developed for the Air Traffic Control (ATC) services can provide to improve traffic management for departing and arriving aircraft. Advanced Surface Movement Guidance and Control Systems (A-SMGCS) have been developed over the years in order to provide controllers with more information (Jakobi et al. 2010). While research often addresses what kind of additional information can be integrated into A-SMCG-Systems, little research discusses factors that limit an air traffic controller's contribution to efficient traffic management at the airports. Hadley et al. (1999) state that although there is a large number of performance measures for air traffic control service the interrelation between controller performance and system effectiveness is not well explored. Both the human operator and system constraints define the performance of the overall system. Still today, there is a clear need to separate system variables limiting the performance of the system from the controller's choices to optimize traffic management at the airport, in order to optimise human-automation systems.

The objective of this paper is to introduce Shannon's entropy H as a metric for the description of multi-aircraft departure time distributions (Shannon & Weaver, 1948). In a second step, this paper tests whether the metric is valuable for the prediction of the system performance metrics of an airport, such as the deviation from on-time departures. Applying new concepts to characterize air traffic at airports and to understand their influence on the system is an important step towards improving the management of the future ATC-system. Such insights might help to answer the questions, when, why and how to assist air traffic controllers to optimize their performance. Further, these insights may make it obvious when controller performance is most critical for the overall system performance and, in contrast, when other system variables decrease and limit the contribution of good controller performance.

A. ATC-Service: The ground control working position

Today, the task of a ground controller, who is responsible for departing traffic, is not only about getting aircraft to the runway as soon as possible, but also about ensuring that there is not a long queue of aircraft back from the holding point, short of the runway. In the complex air traffic system it might be desirable to always have one or two aircraft at the runway ready for departure, as this might serve as a buffer in the system. However, from an environmental perspective it is better to minimize the waiting times of these aircraft having their engines running.

Therefore, assistance systems are designed to assist ground traffic controllers in deciding when to release aircraft for taxiing to the runway (Anagnostakis et al., 2001). This human-automation system (controller and assistance system) still faces a challenging task because there are competing goals and there is not one optimal solution. To meet the obligations, the human-automation system should release aircraft so that they depart on-time. In addition, supported by the assistance system, the controller should minimize taxi delay times and ensure good traffic throughput. Research about new controller assistance systems is intended to ensure the optimal performance of the human-automation system. Therefore, simulation studies are completed to explore the performance of new human-automation systems.

B. Human-in-the loop simulation studies

Scenarios of a human-in-the-loop simulation, exploring ground controllers' working positions, should represent snapshots of controllers' shifts which can be used to estimate or predict how a new human-automation system may perform in the real world. Data are recorded in fixed time interval scenarios (e.g. 20 min) in order to collect a dataset representative of controllers at work. On the one hand it is an advantage that the designer of the simulation study is able to define traffic scenarios, but on the other hand it has to be proven that these snapshots are valid for much larger time intervals in real world operations.

From a research perspective fixed scenarios also reduce complexity and allow for systematic exploration of correlations within the data. From a psychological perspective, data from simulation experiments are still very complex and dependent variables might be influenced by not only controllers' performance, but also by other variables within the complex system.

C. Predicting temporal deviations from on-time departure performance

One system performance criterion of interest when studying controller performance is the on-time departure performance. Controllers persue the goal to minimize the temporal deviations of actual wheels-up times of aircraft from the scheduled departure times. For some aircraft departure schedules this goal might seem more challenging to controllers than for others. Analysing data from simulation studies including different aircraft departure schedules allows for research into which system variables have any explanatory power to predict temporal deviations from on-time departure in a systematic way.

An aircraft departure schedule for one simulation scenario might start with a stack of aircraft which all should be released early in the scenario, while in another scenario there is not much traffic at the beginning, but a stack of aircraft that should depart later. The *mean scheduled departure time* (MSDT), defined as the mean of the scheduled departure times for all aircraft within a given simulation scenario, is one possible variable that might explain on-time departure performance.

However, a variable, such as the MSDT, does not include information about whether the aircraft departure schedule of a scenario is very stacked. To research whether the "stackedness" of the aircraft departure time distribution is a predictor for on-time departure performance another metric is needed. Looking for such a metric, it seemed reasonable that Shannon's entropy might offer a valuable solution.

D. Shannon's entropy H

Shannon's transfer information theory is one of the foundations of information theory. Shannon's theory explains how content of information can be described by entropy in binary units. According to his theory a completely predictable message equates to an entropy value of zero, while less predictable messages show higher entropy values.

In information theory complex systems are defined by entropy. A complex system is one in which uncertainty is high (low entropy), because it can be characterized by a high degree of randomness.

Reviewing the literature of human factor research in ATC there is not a lot of research addressing entropy. Hilburn (2004) reviewed studies using entropy. According to Hilburn, the entropy concept was applied to eye scan behavior by Harris, Glover & Spady (1986). For ATC, Mehadhebi (1996) used entropy to predict traffic arrival locations, while Hansman and Histon (2002) captured the general dispersion of traffic.

Entropy was used in a different context by Belavkin and Ritter (2003) for the analysis and control of cognitive models. Further, Röttger et al. (2007) used entropy to measure orderliness of sequences of control actions. In the following section it will be explained how the entropy concept can account for departure time distributions in departure schedules.

E. Stacked versus uniform aircraft departure schedules

At airports, it is a well-known fact that there are peaks of aircraft departures in the mornings and evenings, increasing controllers' task- and workload. That being the case, it is also a common understanding that departure delays increase during these time periods.

Within these peak times more than one aircraft is scheduled for the same desired departure time. As a result, the aircraft are stacked (statistically, non-uniform distributed). To visualize this idea, compare Figures 1 and 2, where Figure 1 is an example in which six aircraft are stacked for a scheduled departure time at 11:00. Figure 2 is an example in which the same aircraft are more uniform distributed (less stacked) to three different scheduled departure times, resulting in two aircraft being scheduled for 11:00, two for 11:10 and two for 11:20.

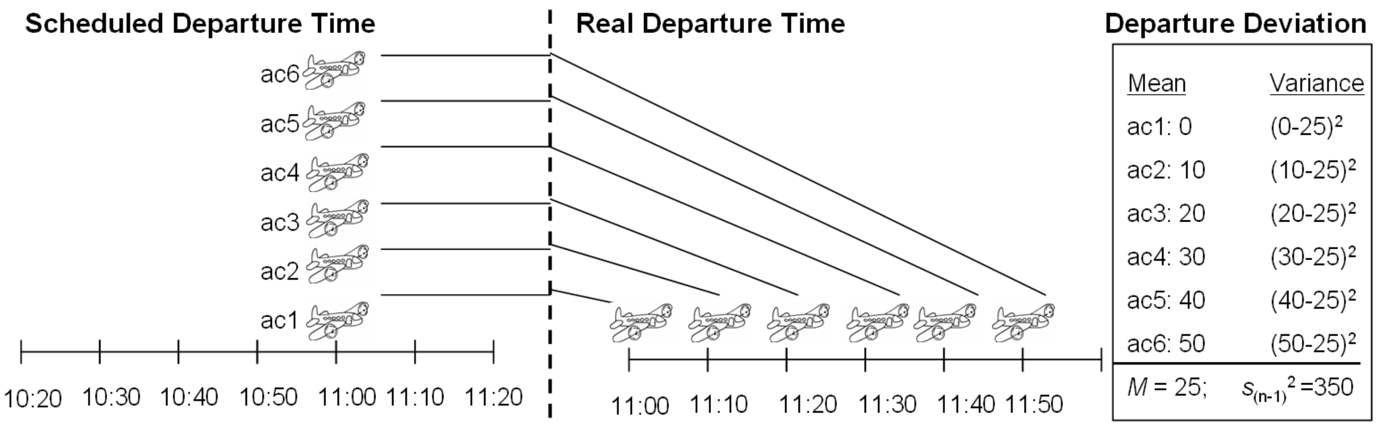

Figure 1. Illustrating "stacked" aircraft (ac) represented by a low entropy value

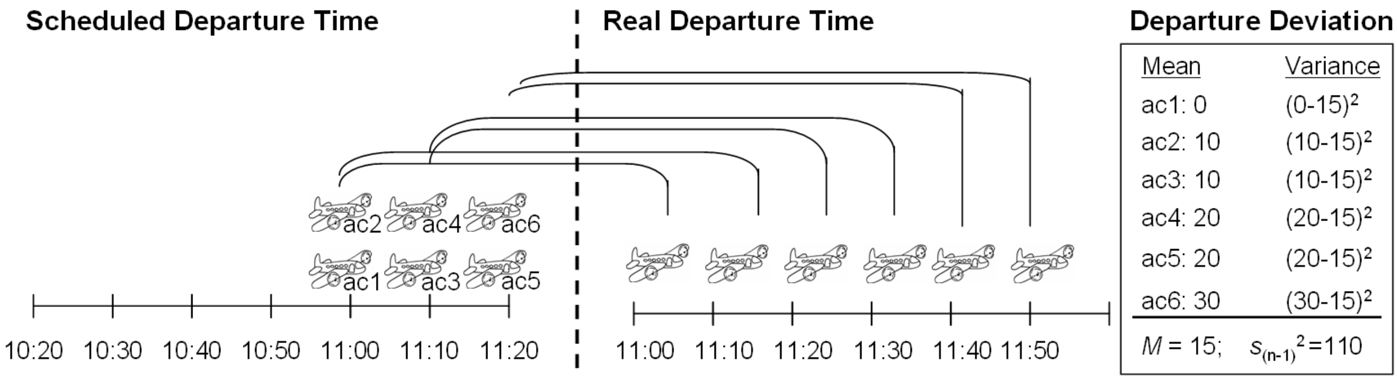

Figure 2. Illustrating "less stacked" aircraft (ac) represented by a high entropy value

From now on we use the metric entropy to characterize this feature of an aircraft departure schedule. A low entropy value represents the fact that aircraft are stacked within a scenario, in contrast to high entropy values, in which scheduled departure times are more distributed over time.

F. Theoretical considerations

In the example (Figures 1 and 2) we assumed that the traffic at the airport was punctual and aircraft took off according to the aircraft departure schedule. When we calculated the temporal deviations from on-time departure (departure deviation time, DDT) for the example, the results showed that the stacked aircraft departure schedule resulted in a higher departure deviation time (Figure 1, right: M=25) than the less stacked schedule (Figure 2, right: M=15).

However, during peak times or due to bad weather conditions, it is likely that the assistance system and the controller will get behind schedule and none of the aircraft can be released on-time. Figures 3-5 exemplify this situation. If the first aircraft is already late (actual departure 11:00), whether there are six aircraft stacked at 10:30 (Figure 3), or there are six aircraft scheduled, two for 10:20, two for 10:30 and two aircraft for 10:40 (Figure 4), both departure schedules would show on average the same deviation from on-time departure.,. It is also interesting to note that irrespective of the order in which aircraft are released, the mean departure deviation time remains the same, even though the variances differ (see Figure 4 and 5). From this second example it might be concluded that entropy cannot explain the deviations from on-time departure. While the stacked (Figure 3) and less stacked (Figure 4) traffic scenario differ in their entropy values, they show on average the same departure deviation times. However, we must look more closely at the correlation between entropy and average departure deviation times. In both cases, the human-automation system is facing a stack of six aircraft at 11.00 (=same entropy), and from this perspective it is plausible that the mean departure deviation times would remain the same.

A similar problem occurs when the human-automation system is able to release aircraft earlier than scheduled. If a controller or the automation system detects that there might be a stack of aircraft later on, s/he might start to unstack the initialized traffic scenario by releasing aircraft early whenever possible. Also, in these cases the characteristic of the initialized traffic scenario is not representative; whereas the traffic scenario as it was modified by the controller is representative for the purpose of calculating correlations between entropy and the temporal deviations from on-time departure performance. These theoretical considerations are important for the input data of the regression models used later on.

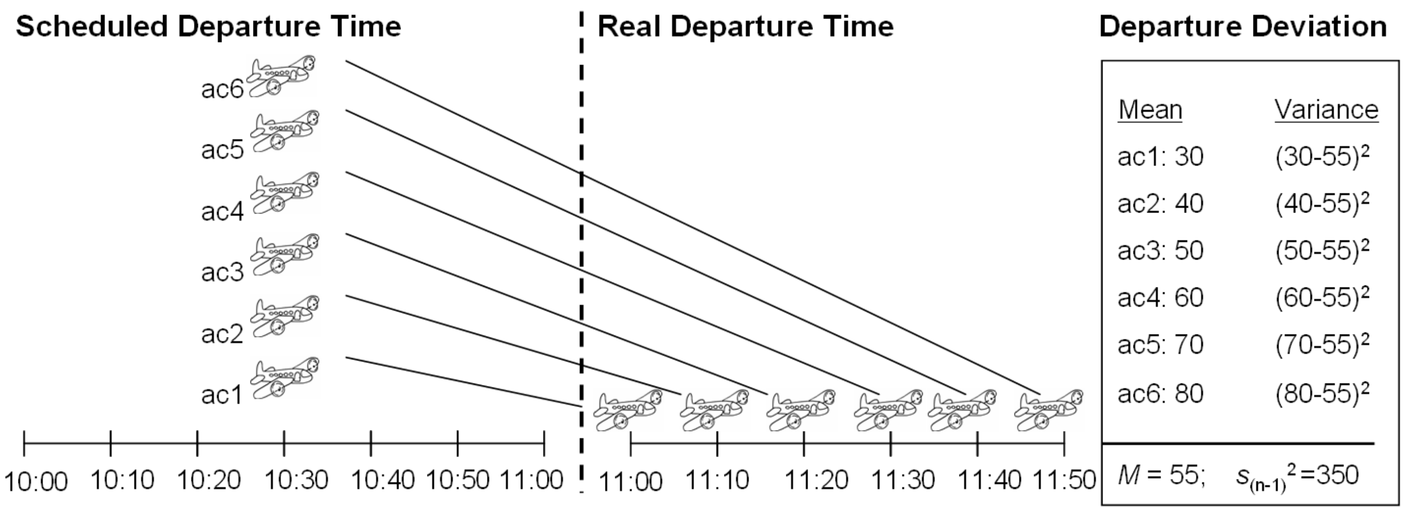

Figure 3. Illustrating six aircraft (ac) stacked and released late: low entropy, $M_{(DDT)}$=55

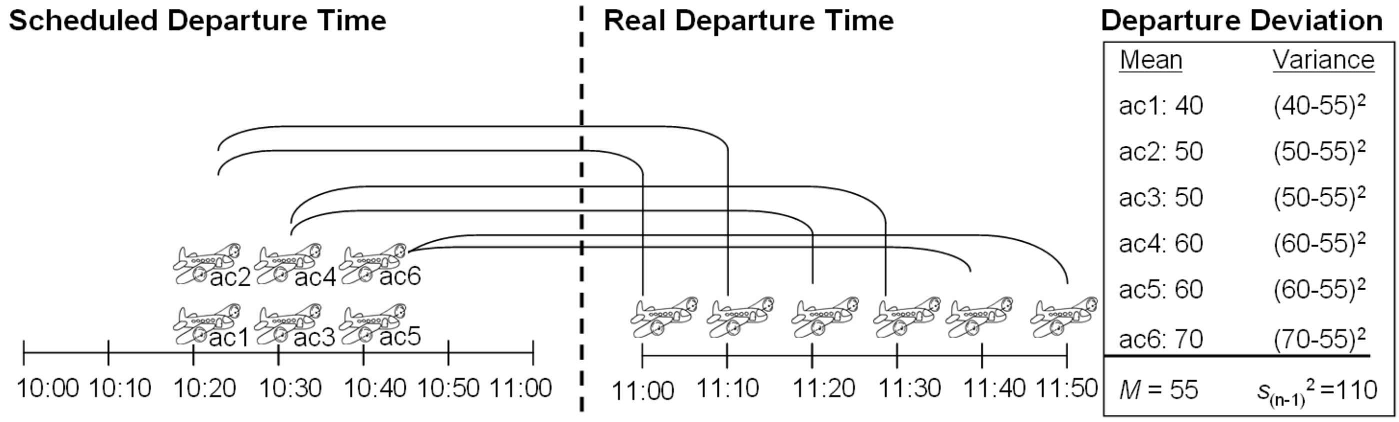

Figure 4. Illustrating less stacked aircraft (ac) released late: higher entropy, $M_{(DDT)}$=55

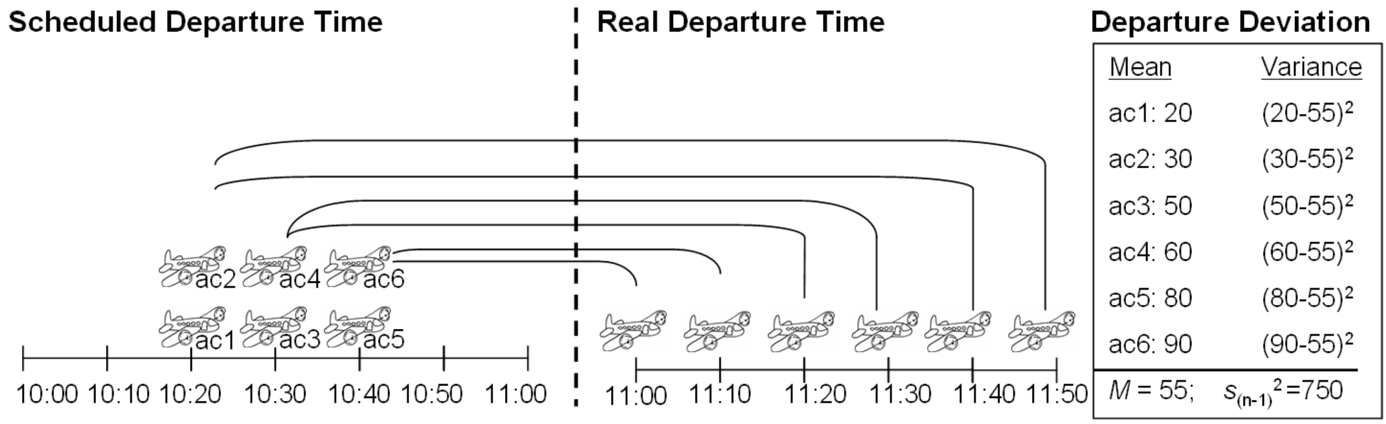

Figure 5. Illustrating less stacked aircraft (ac) released in different order: same $M_{(DDT)}$=55

II. METHOD

A. ATC-simulation study

For the entropy analysis a data set of a medium-fidelity ground controller simulation study, completed at the University of Illinois Urbana-Champaign, is used.

Sample. For the study, 21 pilots, as a convenient sample, were recruited from the University of Illinois, Institute of Aviation. They were all males, between 19 to 25 years old and were paid $20 per hour.

Simulation environment. The study was conducted using a medium-fidelity simulator setup, simulating a 150° degree tower view on the East Terminals of Dallas Fort Worth International Airport (DFW). The different assistance variants for the experimental conditions were integrated into the ground control display, which was placed right in front of the participants. No voice communication was realized. Participants interacted with the system using electronic flight strips via mouse and a X- keypad.

Task. The participants were asked to execute the task of a ground controller at Dallas Fort Worth International Airport. They were responsible for the incoming and outgoing air traffic of the east Terminals. Responsibility for arriving traffic was handed over from the tower at three different spots on the west side of the airport. Traffic departing from the terminals waited for the ground controllers' taxiing instructions and then was handed over to the tower for take off clearances. For completing their ground control task, participants had to decide when to release departing aircraft for taxiing to the runway. They were told to accomplish four goals: (1) Maximize throughput, (2) Depart aircraft on time, (3) Minimize the number of stops to save fuel and (4) Handle arrivals. Participants were instructed how to pursue all of the four goals in parallel in more detail.

Traffic Scenarios. The generation of traffic scenarios was based on real traffic data collected by the SODAA Tool developed from Mosaic ATM (Brinton et al. 2010). This tool allows for re-creating actual operational traffic scenarios for simulator studies to ensure a representative traffic sample of specific airports. 11 scenarios (20 min each) were created for this simulator study based on traffic data recorded from DFW August 17th-23rd 2009.

Experimental Design. A 3*8 between factor design with the factors visualisation aid (Control, TCV, ORT) and traffic scenarios was used. 11 traffic scenarios varied in positioning and timing of arriving and departing aircraft, of which three were used as training scenarios. Participants were assigned randomly to the three different experimental conditions. In the Conventional Traffic Display Condition (Control) participants worked with a conventional traffic display representing ground traffic movements with an update frequency of 1 second. In the Temporal Constraint Visualisation Aid Condition (TCV) additional information is integrated into the graphic display. Bars are visible next to the taxiway to facilitate the controllers' perception of temporal constraints, when to release aircraft for departure. In the Optimal Release Time Condition (ORT) a timeline presents to the controller when s/he should release the departing aircraft. In the experiment an algorithm for departure management was used, not provided with information about arrivals on the taxiways. The algorithm was further restricted in that way that it was not considered, to release aircraft prior to there scheduled departure time, if possible. Recalculations of the algorithm were updated on the display every second.

Procedure. Participants completed sessions of 3 hours on 2 succeeding days. There were three different versions of instructions, one for each of the three experimental conditions. The instructions were followed by the training trials. The different assistance variants for the experimental conditions were integrated into the ground control display, which was placed directly in front of the participants. No voice communication was used. Participants interacted with the system using electronic flight strips via mouse and a X-keypad.

B. *Calculating entropy values for departure schedules*

The departure schedules within the simulation traffic scenarios differed in their numbers of departing aircraft. Participants released at least eleven aircraft (ac) per scenario. For a systematic comparison, therefore only eleven aircraft are included for the following analysis. Entropy values were calculated using the equation for entropy H developed by Shannon. For example, the entropy for scenario 11 (7 bins with 1 ac; 2 bins with 2 ac; 2 bins with 0 ac) was calculated as follows:

$$\text{Entropy } H = 7*(1/11*\text{Ln}(11)/\text{Ln}(2)) + 2*(1/11*\text{Ln}(11/2)/\text{Ln}(2)) + 2*(1/11*\text{Ln}(1)/\text{Ln}(2))$$

Figure 6 depicts the nature of the aircraft departure schedules of the input scenarios used in the simulation study.

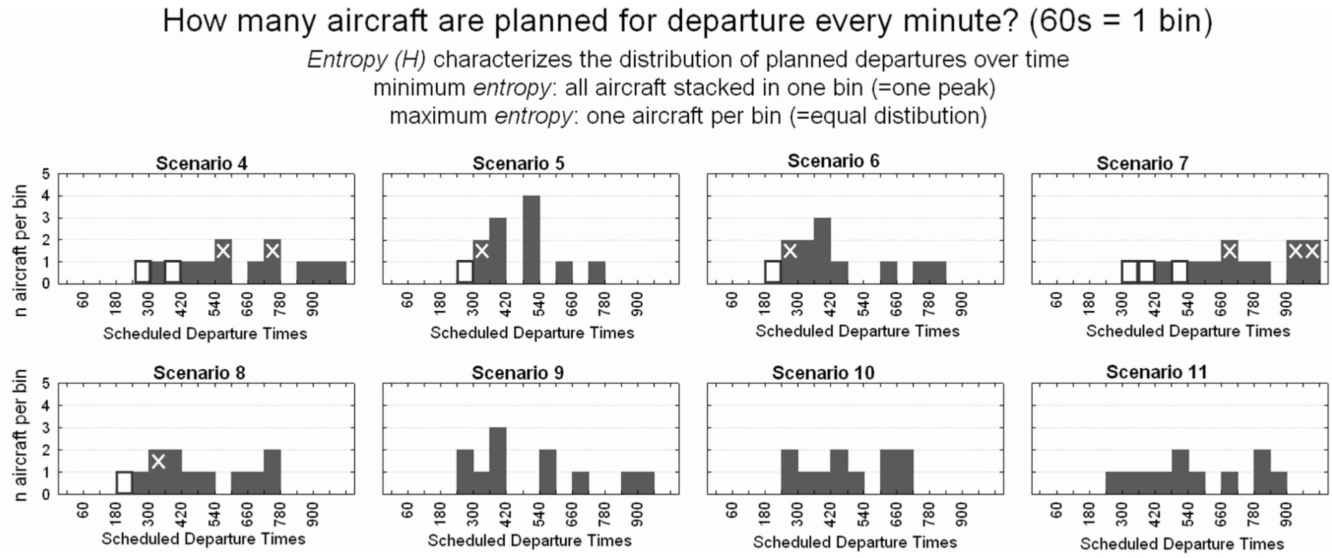

Figure 6. Departure schedules and how they were modified for the control and TCV condition.

The theoretical considerations discussed earlier have shown why it is important to verify whether these input scenarios are representative of how controllers work traffic. Due to the fact that in the Control- and the TCV-Aid Condition the strategy "to release aircraft early" was used, additional entropy values (entropy*) were calculated for traffic scenarios 4-8. The white Xs and empty boxes in Figure 6 show that participants in these conditions unstacked traffic and therefore modified the initial setup.

III. RESULTS

Regression models were calculated for each experimental condition to test whether the stackedness of the scenarios in terms of entropy can predict the differences in scenarios' temporal deviations from on-time departures, or equivalenly their scenario departure deviation times (DDTs). The results of the regression models also indicate what proportion of the variance in DDTs can be explained by entropy.

A. *Entropy to predict Departure Deviation Times*

First, regression models were calculated based on entropy values for the departure schedules as they were initialized. For all conditions entropy significantly predicted DDTs. For the Control Condition the regression coefficient was b=-.66, $t(54)$=-6.39, p<.001, for the Time Constraint Visualization Aid Condition it was b=-.69, $t(54)$=-7.09, p<.001 and for the Optimal Release Time Condition b=-.88, $t(54)$=-13.43, p<.001.

Entropy also explained a significant proportion of variance in scenarios' DDTs within each condition; for the Control Condition with R^2=.43, $F(1,54)$=40.92, p<.001 and for the Time Constraint Visualization Aid Condition with R^2=.48, $F(1,54)$=50.30, p<.001. Among the three conditions entropy explained the highest proportion of variance for the Optimal Release Time Condition with R^2=.77, $F(1,54)$=180.36, p<.001.

B. *Entropy* to predict Departure Deviation Times*

Second, for the Control and Time Constraint Visualisation Aid Conditions we calculated regression models based on entropy* values, which integrate the fact that within these conditions participants used the strategy "to release aircraft early" and in this way modified the initial traffic scenario. As expected, entropy* was a better predictor of scenarios' DDTs than entropy. For the Control Condition the regression coefficient showed a stronger negative correlation b^*=-.75, $t(54)$=-8.35, p<.001 and the proportion of explained variance increased to R^{2*}=.56, $F(1,54)=69.72$, $p<.001$. For the Time Constraint Visualisation Aid Condition the regression coefficient for entropy* also showed a stronger negative correlation $b^*=-.77$, $t(54)=-8.96$, p<.001 and R^{2*}=.60, $F(1,54)=80.40$, $p<.001$, than the correlation between entropy and scenario's DDTs. For the Optimal Release Time Condition, as expected, entropy remained the better predictor in contrast to entropy*, because participants mainly did not release aircraft early. The results $b^*=-.78$, $t(54)=-9.14$, p<.001 and $R^2=.60$, $F(1,54)=83.55$, $p<.001$ confirm this assumption.

C. Summary

As expected, for the Control Condition and Time Constraint Visualisation Aid the variable entropy* was able to improve the prediction of differences in scenario's DDTs. For the Optimal Release Time Condition entropy (based on the initialized departure schedules) remained the better predictor. Therefore, in Figures 7, 8 and 9 the scatterplots and regression lines for each condition are depicted, using entropy* or entropy whichever variable was the better predictor for the condition.

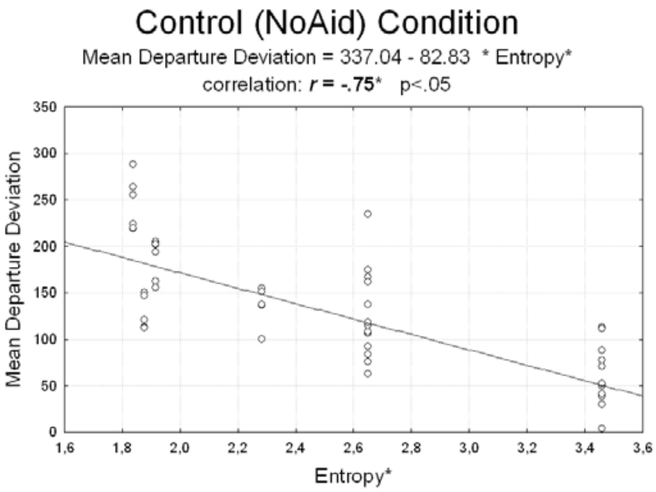

Figure 7. The regression line shows the linear fit (least square) for the variables entropy* and Mean Departure Deviation Time per scenario.

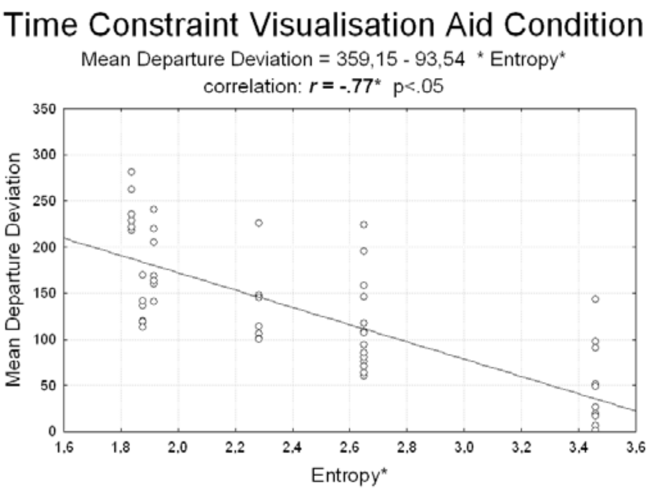

Figure 8. The regression line shows the linear fit (least square) for the variables entropy* and Mean Departure Deviation Time per scenario.

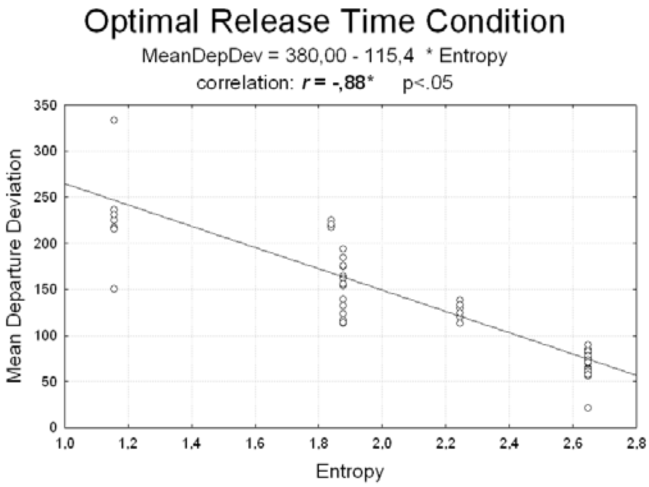

Figure 9. The regression line shows the linear fit (least square) for the variables entropy* and Mean Departure Deviation Time per scenario.

IV. DISCUSSION

A. Entropy to account for stacked departure traffic

Shannon's entropy was applied in this paper to characterize departure schedules in terms of stackedness (statistically, uniform versus non-uniform) to explore whether this metric has explanatory power to predict the departure deviation times (DDTs) for the scenarios. The strong correlations (r*=-.75, r*=-.77 and r= -.88) showed that entropy (or rather entropy*) is a strong predictor for scenarios' DDTs. There are many factors influencing the performance metrics of an airport. The data analysis completed in this paper shows that there is a strong correlation between stackedness of the departure schedule and DDTs, to some degree independent of the experimental conditions under which the traffic scenarios were completed. This indicates that the performance metric DDT is strongly influenced by the system variable for stackedness of traffic and is less well explained by differences in controller performance. Nevertheless, the regression model was sensitive to different control strategies of the human-automation system. If the human-automation system unstacked traffic by releasing aircraft early, the strength of the correlation decreased. However, it was also shown that in these cases the stackedness of the reorganized aircraft departure schedule (entropy*) became an even stronger predictor for DDTs than the original entropy variable.

B. Strength and Limitations of the entropy metric

The strength of the metric entropy is that it accounts for a structural component of the aircraft departure schedule. A metric, such as the *mean scheduled departure time* (MSDT) does not include information, if multiple aircraft are scheduled for the same departure time, or if their departure times are more distributed. However, the entropy metric includes this information. Analyzing how stacked departure traffic is scheduled might also be a sensitive metric to decide about control strategies. The degree of stackedness might indicate when to use a certain control strategy so that the human-automation system completes departure traffic most efficiently. However, the generalization of these results to operational data at airports must be tested thoroughly.

In this paper we focused on a systematic analysis of the metric entropy. It must be noted that there might be other system variables or metrics that can explain the differences in DDTs among scenarios. In addition, while entropy H can be calculated for any number of aircraft in a scenario, we were able to show these strong correlations to predict DDTs only when the number of aircraft in each scenario was equalized, based on the minimum number of aircraft released by participants.

The entropy metric can also predict a mean departure deviation time only for departure schedules based on multiple aircraft within a traffic scenario. It is not able to predict DDTs of individual aircraft.

C. Controllers' performances in human-automation systems

Within the experiment two different assistance visualization aids were implemented. In the Control Condition and the Time Constraint Visualization Aid Condition participants used a strategy "releasing aircraft early", which was not used in the Optimal Release Time Condition. Such a strategy influences the performance of a human-automation system before peak times, but is not of major interest when departing traffic already exceeds the runway capacity.

The regression models indicate that besides this control strategy, the scenario differences in DDTs are mainly due to stackedness of the initial traffic scenario and not to controller performance or rather the performance of the human-automation system. Understanding factors like the stackedness of traffic is therefore important to determine whether a performance metric is sensitive to controller performance or is mainly caused by a system variable. Other system performance metrics in contrast to the DDT might be best predicted by differences in controller performance and less by system variables.

D. Environmental Constraints

Not much research has explored the relationship between controller performance and system performance metrics (Hadley et al. 1999). We used a clean dataset from a research simulation study to research whether an entropy-based approach can provide a quantitative measure of aircraft departure schedules to predict participants' performances in DDTs. In this paper it was shown that the environmental constraints, characterized by entropy have a much stronger influence on the system performance metric DDT than the controllers' performances. To predict human performances in complex systems a good model for the description of the task environment is necessary (Kirlik 93).

Describing the environment of the ground controller using entropy values for aircraft departure schedules seems a promising approach, as it also can account for the strategy "releasing aircraft early", used by the participants completing the ATC- ground control task.

V. CONCLUSION

Understanding which system variables contribute the most to system performance metrics is important in order to conclude whether system constraints or human performance can explain the variability in system performance. It was shown that Entropy H functioned as a good predictor for DDTs within the data set used in this study. Exploring which quantitative metrics, such as entropy H, can explain variance of system performance is important in order to improve the future air traffic management system. However, for a generalisation of these results to operational data, further research is needed. Using Entropy appears promising in order to distinguish controller performance, or rather the performance of the human-automation system, from other influencing system variables. Especially for complex systems, such as an airport, these distinctions are important in order to evaluate the performance of novel human-automation systems.

ACKNOWLEDGMENT

This research was supported by a grant from NASA Ames Research Center to Aptima, Inc. and the University of Illinois Urbana-Champaign. The authors thank Ron Carbonari and Kenyon Riddle for setting up and completing the simulation study this research was based on. Most of the issues presented in this paper resulted from gainful discussions with Ron Carbonari and the research group at UIUC, including Donald Talleur, Ryan Malawy, Reggie Tabachnick and Arjun Rao.

REFERENCES

[1] Anagnostakis, I., Clarke J.-P., Boehme, D., Voelckers U. (2001) Runway Operations Planning and Control: Sequencing and Scheduling. *Journal of Aircraft,* Vol. 38 (6).

[2] Belavkin, R. V. and Ritter, F. E. (2003). The Use of Entropy for Analysis and Control of Cognitive Models. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modelling.* Bamberg, Germany

[3] Brinton, C., Lindsey, J., & Graham, M. (2010). The Surface Operations Data Analysis and Adaptation tool: Innovations and applications. Proceedings of the 29th *Digital Avionics Systems Conference*.

[4] Christien, R., Benkouar, A., Chaboud, T. & Loubieres, P. (2003). Air traffic complexity indicators & ATC sectors classification. Presended at the ATM Seminar, Budapest, 2003.

[5] Hadley, G. A., Guttman, J. A. & Stringer, P. G. (1999) Air traffic control specialist performance measurement database *(DOT/FAA/CT-TN99/17)*. Atlantic City International Airport: Federal Aviation Administration William J. Hughes Technical Center.

[6] Harris, R.L., Glover, B.L., & Spady, A.A. (1986). Analytical Techniques of Pilot Scanning Behavior and Their Application (NASA Technical Paper 2525).

[7] Hilburn B. (2004) Cognitive Complexity in Air Traffic Control: A Literature Review, EEC Note No. 04/04.

[8] Histon, J., Aigoin, G., Delahaye, D., Hansman, J., and Puechmorel, S. (2002) Introducing structural considerations into complexity metrics, Air Traffic Control Quarterly, Vol. 10(2), 2002.

[9] Jakobi, J., Teotino, D. and Montebello, P. (2010). Towards Higher-Level Services of Advanced Surface Movement Guidance and Control Systems (A-SMGCS). *Air Traffic Control Quarterly*, Vol. 18(21)

[10] Kirlik (1993) Modelling strategic behaviour in human-automation interaction - Why an aid can (and should) go unused, Human Factors, Vol. 35(2)

[11] Shannon, C. E., & Weaver, W. (1948). The mathematical theory of communication. Urbana: University of Illinois Press.