Predicting Arrival Time Uncertainty from Actual Flight Information

M. Tielrooij, C. Borst, M. M. van Paassen, M. Mulder
Section Control & Simulation Faculty of Aerospace Engineering
Delft University of Technology
Delft, The Netherlands
m.tielrooij@tudelft.nl

Abstract—**Air traffic control uses Arrival Managers (AMANs) to schedule an inbound stream of aircraft. As these systems use predicted arrival times to optimize the planning for capacity and flight efficiency, the accuracy of these predictions is an important parameter in arrival planning. While prediction capabilities have improved, and are likely to improve more, it is unlikely that prediction error will disappear altogether. Especially in future scenarios with longer planning horizons, techniques will have to be found to support planning in the presence of prediction uncertainty. To enable working with uncertainty on a predicted arrival time, that uncertainty needs to be predicted itself. This paper proposes and tests a method to predict arrival time uncertainty based on historic prediction accuracy using currently available arrival time estimates.**

Keywords- Air Traffic Control, Arrival Manager, Arrival Time Uncertainty, Queue Management, Trajectory Prediction

I. INTRODUCTION

Many Air Navigation Service Providers (ANSPs) nowadays use AMANs to plan the arrival times of inbound aircraft and thereby balance the demand to the available capacity. These systems provide support to the sequence manager (US: Traffic Management Coordinator) in deciding how to best modify the arrival time of inbound aircraft when these are predicted to arrive with too little spacing between them. When aircraft are assumed to fly the optimal trajectory from an Airspace User (AU)'s perspective, deviations in the 4D path from these trajectories should be kept to a minimum. However when deviations are required, earlier decisions allow for smaller deviations, which increases flight efficiency. For example, a smaller speed increase over a longer flight time is more fuel efficient than a larger speed increase over a shorter time, while achieving the same difference in time.

Currently, the horizons at which AMAN are used to monitor and influence traffic are limited by three factors:

1) the availability of information on the predicted arrival time of aircraft, for example the limit of radar surveillance,

2) the authority to influence the aircraft, for example beyond the boundaries of Flight Information Regions (FIRs), and,

3) the reliability of the predicted arrival times.

These factors currently limit the use of AMAN to a horizon of typically 20 to 30 minutes (or 150-200 NM) [1]. Future operational concepts, such as those proposed in Single European Sky ATM Research (SESAR) and Next Generation Air Traffic Management System (NextGen) foresee an increase in the planning horizon to increase the long-term efficiency and predictability of operations. Depending on the concept, the future horizons are expected to be at 200-500 NM, or about 2 hours of flying time [2], [3]. To achieve these increases, the three limitations on the current AMAN horizon need to be overcome.

The first two of these constraints are being addressed in current developments: System Wide Information Management (SWIM) is foreseen to enable continuous sharing of all relevant information concerning a flight between all involved actors [4], [5]. And, through SWIM, different ANSPs are expected to be able to share their requirements on a trajectory (such as an arrival time planned by AMAN) as well as their capabilities to provide for such requirements. While these developments resolve the first two limitations on the planning horizon, the third problem – prediction uncertainty – is expected to reduce, but is unlikely to disappear altogether. Since uncertainty is not expected to be eliminated, increasing the AMAN horizon will require ways to perform arrival planning in the presence of uncertainty.

Any such process that make the uncertainty part of the decision making process will require a sufficiently accurate estimate of that uncertainty. This paper will propose a method to predict the uncertainty on an Estimated Time of Arrival (ETA) based on readily available information on flight progress.

The next section discusses current methods to predict uncertainty. Section III will then analyse the prediction errors in Eurocontrol's Flight Update Message (FUM) messages that ANSPs currently use to report flight progress. This analyses will use flight data for Amsterdam Aircraft Schiphol (AAS) in 2013. Based on the analysis, and the different modelling techniques, Section IV will then describe a method to predict uncertainty by estimating distributions using historic data. The method is validated in Section V. Finally, Section VI will discuss the results of this validation and the applicability of the method to other airports and to other data sources.

II. A REVIEW ON MODELLING ARRIVAL TIME UNCERTAINTY

The accuracy of arrival time predictions strongly depends on the remaining time to arrival. The foreseen arrival management process starts at 2 hours before landing, and continues up to typically 10 minutes before landing when tactical Air Traffic Control (ATC) takes over. Research on prediction accuracy has been performed for both strategic flow management as well as tactical Medium Term Conflict Detection (MTCD). From this research, the different causes of prediction error and their relative effect on the prediction uncertainty with respect to the horizon may be described.

A. Prediction Accuracy

Current operational prediction capabilities depends strongly on the phase of flight. Typical standard deviations for airborne flights are 30 seconds at 20 minutes before an arrival point when airborne, e.g., Flight Management System (FMS) predictions as analysed by Bronsvoort [6]. These standard deviations increase to 15 minutes when the aircraft is still on the ground, e.g., the departure accuracies of several airports in the US found by Mueller and Chatterji [7].

Prior research on prediction of uncertainty in the Air Traffic Management (ATM) domain typically focuses on short horizons (i.e., 20 minutes) for tactical tools, or very long horizons (i.e., multiple hours) for flow management and strategic purposes [8]–[11]. For AMAN, a continuous prediction will be required that provides an estimate of the uncertainty in arrival time, starting at a horizon of two hours until shortly before landing.

At short horizons, the aircraft is airborne, and its current, up-to-date, state is available through surveillance such as radar and Automatic Dependent Surveillance-Broadcast (ADS-B). A study of FMS prediction, shows a typical standard deviation of 60 seconds at 30 minutes before arrival, and 120 seconds at 60 minutes to arrival [6]. Ground based prediction systems are likely to have less information on the actual state of the
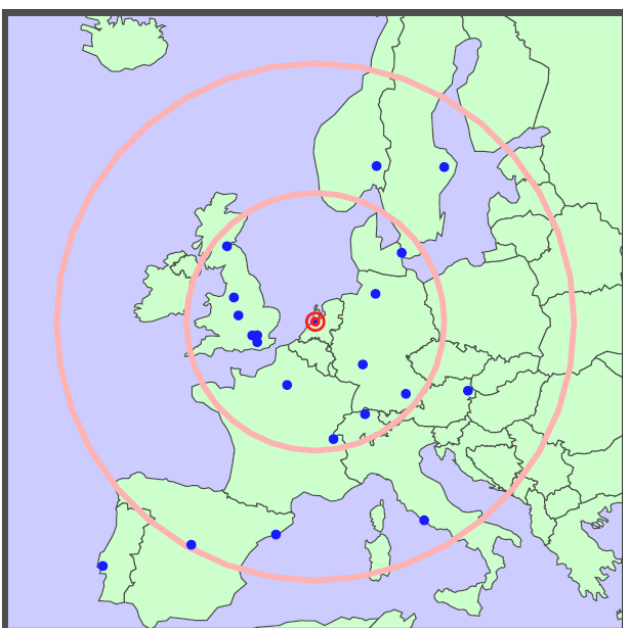

Figure 1. The 20 busiest connections to Amsterdam Airport Schiphol in 2013 according to [13]. The range rings indicate 400 and 800 NM distance, or 1 and 2 hours flight time, respectively.

aircraft, and therefore increase the potential for errors [12]. Similarly, however, the airborne system may be unaware of future ATC decisions that affect its trajectory.

For European airports, many of their busiest connections lie within a 2 hour flight horizon. For example, the 20 busiest connections to Amsterdam Airport Schiphol are within this horizon (Fig. 1) [13]. When the prediction horizon is more than the flight time, the errors associated with taxiing, boarding, and the previous rotation of the aircraft are added to the set of disturbances. The proximity of departure airports makes flight status an important factor in determining prediction uncertainty. The effects of flight status on prediction error have been demonstrated by, for example, research of Solveling [14], and Tobaruela [11].

Since a future system should provide support in continuous decision-making during the tactical phase, the uncertainty associated with an ETA needs to be updated continuously as well. This will require an approach that can be executed on-line during operation[14]. Concluding, a method of predicting uncertainty is required that allows for fast calculation while at the same time modelling the different uncertainties in the different phases of a flight.

B. Predicting Uncertainty

To predict the uncertainty for a particular flight, the iFACTS and CARE projects employ methods based on calculation using propagation of the uncertainty of the input components [8, 9]. This allows for a detailed, and fast prediction over the tactical horizon of approximately 20 minutes. To use this approach in the AMAN context, the added sources of error at longer horizons would require knowledge on the uncertainty associated with those errors. This would both require a far more complicated algorithm, and elaborate research in all contributing components. At longer horizons, this would become impracticable.

To support research into the effects of airspace changes, a monte-carlo approach has been developed that uses empirical information on the outcomes of predictions to model this uncertainty [15]. The authors suggest that such an approach could possibly be used in tactical operation given the availability of sufficient computing power but is currently unsuitable for real-time applications.

It should be noted that, in both described techniques, the uncertainty is calculated for the 3D path as well as the flight time. Since AMAN is used to form an arrival planning based on the arrival time, only the flight time is of importance. At the same time, ANSPs nowadays receive regular updates on the ETA of a flight, both when on the ground as well as in the air. In future scenarios, this capability is expected to be expanded with the advent of SWIM and the Business Trajectory concept. The focus on arrival time only and the availability of ETA estimates provides a means to develop a method to predict uncertainty.

C. Using Empirical Information

Tobaruella et al. [11] used an experience-based approach to predict sector occupancy given predictor uncertainty. This

method estimated errors on ETFMS Flight Data (EFD) messages provided by the Network Manager Operations Centre (NMOC) using historical error patterns. Such a method directly relates prediction error to the actual information available on the flight and the properties of the prediction in particular. This will allow for an uncertainty predictor sensitive to all relevant aspects and fast enough for real-time application. If a distribution can be captured in a small set of parameters, a look-up table with those parameters, based on information in the arrival time prediction, may provide a fast way of predicting error distributions.

The analysis in this paper is based on AAS. Since EFD data is only recently made available, Air Traffic Control the Netherlands (LVNL) currently does not have a large recording of such data, which precludes an analysis of prediction error. However, the key information that is provided in EFD messages is also available in FUMs. The FUM is a message that provides downstream ANSPs with information on the expected arrival time of a flight at its airspace boundary, therefore supporting planning [16]. These messages are currently provided by the NMOC over the Aeronautical Fixed Telecommunication Network (AFTN) and are therefore available. Through using this available information, the technique is readily applicable in current operations.

The FUM is sent whenever one of three conditions occur: First of all, an update is provided three hours before predicted landing. Secondly, the NMOC monitors progress of the flight with respect to their own prediction, if the flight deviates more than 5 minutes, an update is broadcasted. Finally, updates are provided at important changes of the flight status, such as when the aircraft takes off [16].

III. ANALYSIS OF FUM ESTIMATES

In this research, all data for the year 2013 for AAS have been used. A full log of all received AFTN messages has been parsed and filtered for relevant messages on inbound flights. These messages were correlated to a flight based on the callsign, date of flight, departure airport, and arrival airport, and stored in a database.
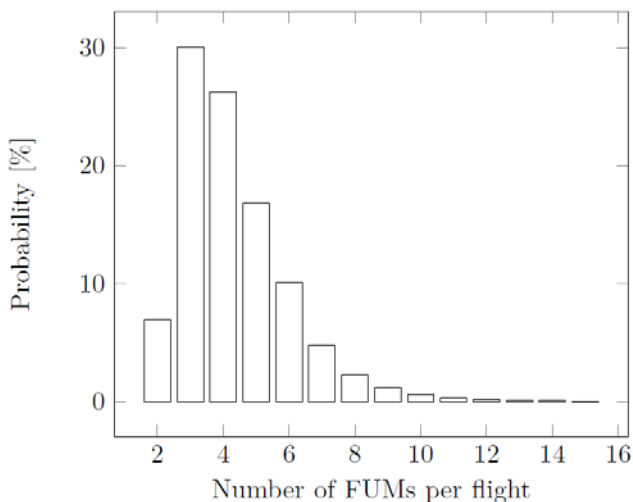

Figure 2. Distributions of the number of messages per flight.

A. Filtering

Initial filtering of the messages applied the following rules to reduce erroneous, or ambiguous data:

- Messages should be formatted according to the standardized formats.

- Messages should contain all elements required for association to a particular flight (callsign, departure date, origin, destination).

- A flight plan message, containing essential information on the flight must have been received before any other messages on that flight are accepted.

- Flights that have been cancelled or diverted are rejected.

- Flights that have errors of more than 12 hours are removed, as analysis showed that these are mostly incorrectly correlated flights of the following day using the same callsign.

- Five days with excessive disruption (due to wind, fog, or industrial action) where removed as the planning process is altered to accommodate the uncertainty on such days. These days were determined by comparing the mean and standard deviation of delay with respect to the scheduled arrival time to weather reports and posterior reports on disruptive events published by the NMOC [20].

The removal of accidentally duplicated flights removed 9,707 flight records; Removal of disruptive events lead to a removal of 4,106 flights. For the remaining 195,759 flights, 657,093 FUMs were received. The number of messages per flight varied around a median of 3 (Fig. 2).

B. Determining Error Distributions

To determine the prediction error for each message, the Estimated Landing Date and Time (ELDT) of the last received message of a flight was assumed as truth data for that flight. In all cases, these messages where termination messages with which the NMOC informs the ANSPs that the flight has been assumed as landed. This message is provided after a confirmation of arrival has been received from the destination airport, or 20 minutes after the ELDT [19]. In the first case, the actual landing time is known; in the second case, it is assumed to be equal to the last ELDT. Since the FUM is only updated when the prediction error is found to be larger than five minutes, the truth data have an inherent error potential of five minutes.

Using the truth data, the prediction error for each message can be calculated. Subtracting the message timestamp from the landing time, each message can also be assigned the actual value for the remaining time before arrival.

To determine the error distribution, the errors of comparable messages for different flights have to be grouped. In the available data, the following properties for a message are available and have been considered for grouping:

- Flight callsign: Each flight may have particular objectives and therefore punctuality requirements. However, flights are only operated once a day at maximum, and would therefore provide a very low number of samples per group.

- Flight operator: Similar to the callsign, punctuality requirements may depend on the AU's business model. Again however, some operators only have one flight per day, and therefore a small sample set.

- Flight status: The flight status indicator shows progress of the flight from filing the flightplan up to termination of the flight. Only the four most prevalent have been considered: When a flightplan is filed (indicated by the acronym FI), when the NMOC has applied a regulation and set a slot time (SI), when NMOC assumed the flight to have departed (i.e., the estimated departure time passed) (TA), and when the flight is reported airborne by ATC (AA).

- Origin airport: Variation in airport operating procedures may well influence punctuality. More importantly, the origin airport determines length of the flight, and therefore whether an aircraft may still be on the ground within the horizon, thus generating the associated additional uncertainty. However, this would be a large subset for the most frequent connections only. Also, the flight status indicator provides similar information on the progress of the flight.

- Date of flight: The date of the flight may account for seasonal disruptive weather effects, such as de-icing during winter for example.

- Planned date and time of arrival: The arrival date provides a similar indicator as the departure date. The arrival time may show daily effects due to airport and airspace congestion.

- Aircraft type: Different aircraft may be more flexible in achieving punctuality; similarly, modern flight systems may provide more support in punctuality. Again however, only a small number of types would generate most of the traffic. Other types would therefore have a small sample.

- Remaining time before arrival: As already indicated, the prediction horizon is the most important parameter in prediction uncertainty.

Selection of the criteria and the width of resolution for those criteria (bin width) is a compromise between detailing of the model and availability of data. Based on an initial analysis of the available information, the following grouping criteria were selected: Horizon (in bins of 10 minutes), flight status, and arrival time of day (in bins of 3 hours).

C. Interpolating errors

Fig. 3 shows a distribution of the messages over the (predicted and actual) remaining time to arrival. The graph shows how, toward the horizon, more and more airborne reports are provided and less messages with a flight status on the ground. The graphs also show the peak in updates at 3 hours before expected arrival, which represents the publishing of the flight plan by NMOC. This peak is much more distributed when plotted against the actual time of arrival due to the prediction errors (compare Fig. 3a to Fig. 3b).
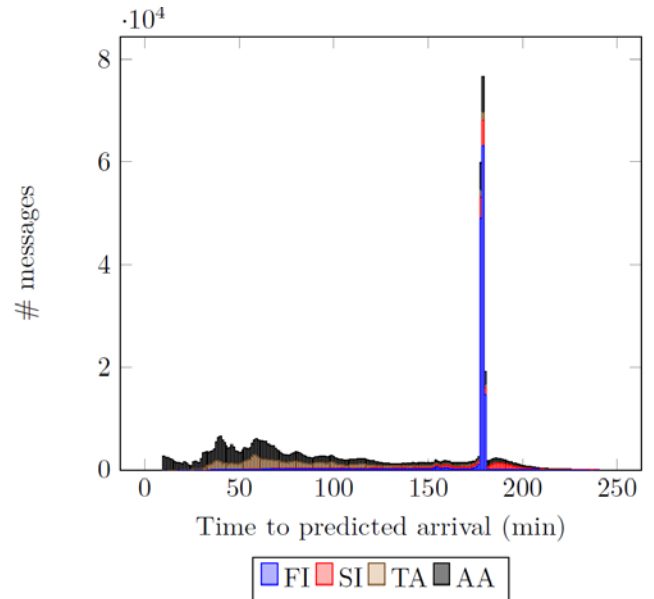
(a) Message count with respect to predicted time to fly
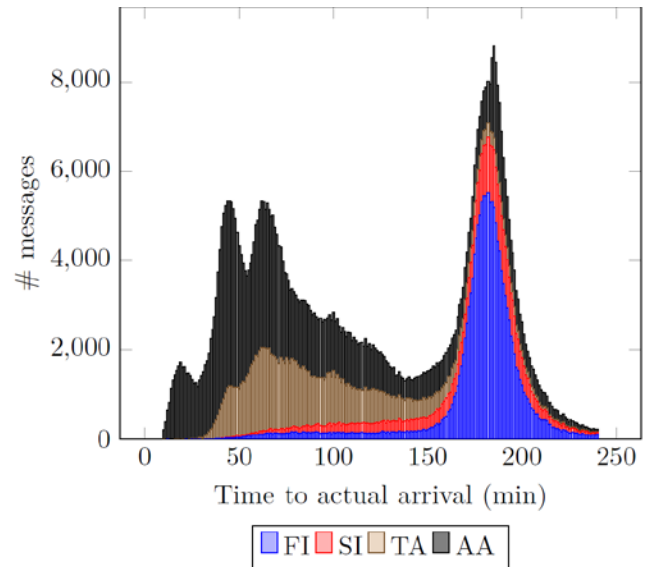
(b) Message count with respect to actual time to fly

Fig. 3. Distribution of messages and different flight statuses over the prediction horizon. The acronyms indicate the flight status: FI – Filed, SI – Slot Issued, TA – Assumed Departed, AA – Confirmed Airborne)

The histograms indicate that the number of messages varies by prediction horizon. Since the NMOC only provides an update when the ETA deviates more than 5 minutes from the expected plan, accurately predicted flights are actually under-reported. Determining the accuracy based on the messages at a certain horizon would overestimate the uncertainty, especially

at horizons close to arrival. To more accurately represent the more accurate flights, the errors need to be interpolated.

When aircraft are airborne, it is likely that the error gradually reduces, as less deviation is possible in the smaller remaining flight time. However, a flight status change is a discrete event, and interpolation between these changes is not straight-forward. Delaying events on the ground are more likely to be discrete, such as for example a report of a missing passenger.

In this work, the errors for reports of airborne flights are assumed to be decreasing over time, and are therefore linearly interpolated. Errors for messages with other flight statuses are assumed to be constant up to the next report. The following analysis is performed on the interpolated values.

D. Prediction Accuracy

Fig. 4 shows the accuracy of the messages versus actual remaining time to fly and flight status. In general, the predictions are biased toward delay. The bias is smallest for the predicted arrival time in the filed flightplan (FI) and particularly strong when the NMOC assumes the aircraft to have departed, but no radar-based report has been received (TA). Finally, when radar reports come in, the predicted time starts approaching the actual time.

The small bias in the mean error of the flightplan suggests that AUs tend to try and achieve their scheduled arrival time even when departing later than scheduled. While this behaviour is seen at other airports, the effect may be exacerbated by the relatively high amount of transfer passengers at AAS, combined with the most prevalent AU (KLM) that operates a hub-and-spoke network from this airport.

The spread of the prediction errors for the filed flightplans appears to be fairly constant, confirming that deviations mainly primarily depend on events occurring when the aircraft is on the ground. Such events do not depend on the time to fly and will therefore not affect the spread of the error with respect to the remaining time. In the other flight statuses, the spread decreases for smaller prediction horizon, with airborne reports showing considerably more precision than other predictions.

Finally, the accuracy and precision at 20 minutes for the non-flying status (FI and SI) is far less than any prediction at larger horizons. These reports may come from local flights, which are cleared for departure in direct coordination with the arrival controller. More likely, they constitute flights for which no progress information was made available to the NMOC, or data were erroneous for other reasons.

Fig. 4 provides the information with respect to the actual arrival time, which has been determined posterior. However, in the operational context, only *predicted* arrival time is available, any estimate of uncertainty will need to be determined on that predicted value. Fig. 5 provides the same information, but now plotted against the predicted arrival time.

The first notable observation is the enlarged deviation of the non-airborne prediction just before arrival. Secondly, the decreasing spread seen in the SI, and TA status in Fig. 4 is less pronounced. Both are an effect of erroneous predictions being distributed over the prediction horizon (i.e., the error of a flight with a prediction error of 20 minutes at 40 minutes before actual landing is counted at 20 minutes before predicted landing). The graph shows that the remaining time to fly – by itself – provides very little information on prediction error.

Air traffic during daytime is considerably more busy than during night time. During daytime demand for departure, en-route and, at the arrival airport regularly reaches available capacity. Departure capacity limits cause delays at the departure airport, en-route, and arrival. Demand is balanced by applying restrictions through the NMOC. These delays affect the accuracy of the expected arrival time, in particular before departure. Fig. 6 shows this effect, as the spread of predictions for flights that have not been confirmed airborne (FI, SI, and TA) increases considerably for flights predicted to arrive during peak traffic.

E. Shape of the Distributions

Figs. 5 and 6 show the median and the spread of the data. However, due to the high number of data points, the exact shape of the distribution is not very clear. To better understand the effects of different flight parameters on the actual distribution, Figs. 7 and 8 plot the distributions of the errors as function of the different properties for a subset of these measurements.

The precision of predictions, and therefore small spread for airborne flights (AA, seen Fig. 5) is confirmed in Fig. 7: The spread of the data is smaller than 20 minutes. For the other flight statuses, the spread is in the order of 1 hour.

When not yet airborne, a skew towards delay (left) is visible. Explanation for this error lies in the fact that aircraft have a maximum speed, but more importantly, passenger and cargo considerations make it impractical to depart before the scheduled departure time.

As the prediction horizon increases, the spread of the error for flights that are airborne (AA), or assumed to be airborne (TA), increases as well, as was visible in Figure 5. The error of the filed flightplan (FI) does not depend on the prediction horizon, except for very short horizons. Here a limited sample set of likely highly erroneous flights, give a false suggestion of inaccuracy.

When compared to the scheduled time of arrival in Fig.8, the number of flightplans scheduled to arrive in the middle of the night with a flight time of less than 60 minutes is very small. The lack of data does not warrant analysis of non-airborne flights within this group. The accuracy of the airborne flights at night is very high, most likely because low traffic levels allow flights to follow their filed route since ATC will not need to vector them to provided sufficient spacing.

Especially during the morning peak, the error spreads considerably. Flights which are assumed to be airborne without a confirmation from an ANSP show an increased skew.

IV. MODELLING DISTRIBUTIONS

One way to develop a fast modelling technique is to make the model as simple as possible. For error distributions, such a

simple model is a normal distribution. Figs. 7 and 8 show a normal distribution based on the data in each sample. The graphs show that the normal distribution only provides a reasonable fit for flights that have been confirmed to be airborne (The bottom row).

The variety of different shapes in the distributions makes the versatile Johnson distribution a suitable candidate for the dataset [18]. This distribution actually consist of a flexible set of 4 distributions which in turn are transformations of the standard normal distribution. The resulting curve can describe any distribution regardless of mean, standard deviation, skew, or kurtosis.

The Johnson-curves can be defined by five parameters. When those parameters have been estimated for different conditions, a hypothetical table-based system could be developed which provides these parameters as a function of the properties of a received prediction. Subsequently the distribution could be generated from the parameters.

A. Fitting Data

By applying the algorithm developed by Hill et. al. [19] a Johnson curve was fitted to each dataset with at least 100 samples using moments [20]. For the available datasets, only the unbounded (SU) and bounded (SB) variants were fitted, with the unbounded variant being the larger majority. The fitted curves are shown in Figs. 7 and 8. The graphs show that the fitted distributions are close to the histograms of the dataset.

Subsequently, the goodness-of-fit was measured using the Kolmogorov-Smirnov statistic (i.e., the maximum absolute difference between the Cumulative Density Functions (CDFs)). However, the Kolmogorov-Smirnov test assumes independence of the two distributions under consideration, which is clearly not true for the fitted distribution. While the test value serves as an indicator of goodness-of- fit, it cannot be compared to a critical value to test significance.

To determine the significance level of the difference between the CDFs, the Kolmogorov-Smirnov statistic was determined by applying a bootstrap resample process. From each subset, a random sample of 1,000 points was selected with replacement. From this sample, the Kolmogorov-Smirnov statistic with respect to the CDF of the Johnson distribution was calculated. The process was then repeated 10000 times to determine the confidence interval for the statistic. Typically, the 95% confidence interval of the statistic reaches a maximum around 2.0 with 1.95 being the cut-off value to statistically reject the hypothesis that the distributions match at a 0.001 confidence level. An exception is flight status AA at a 20 minute horizon where the parameter can be up to 15. This is most likely due to the fact that the discretisation is relatively large compared to the width of the distribution.

The statistical tests show that the fitted function is often statistically different from the dataset. However, the difference in the CDF, which is not multiplied by the square root of the number of sample points, will be in the order of 1%. Making decisions based on the CDF likely to be accurate.

V. VALIDATION

Since the Johnson distributions have algebraic equations for both the Probability Density Function (PDF) and the CDF, a look-up table with the parameters provides a fast way of calculating these distributions as needed. The suitability of the method then depends on the robustness of the fitting technique, and the effects of the discrete bins on the continuously changing distributions.

To test the robustness of this technique, a process akin to the bootstrap process was applied in testing the goodness-of-fit of an estimated Johnson curve. A subsample of 4/5th of the dataset was selected at random and used to fit a Johnson-curve. Subsequently, goodness-of-fit was determined with respect to the complementary sub-sample. This process was repeated 50 times to determine the confidence bounds for each subset of data. The (unscaled) Kolmogorov-Smirnov statistic varies between up to 0.02 for the FI and AA flight statuses, and 0.1 for SI and TA . This shows that, while sampling has an effect, the effect on the CDF is unlikely to be large.

If uncertainty is to be used in an on-line display, gradual changes are preferable to discrete changes. While the change of a flight status will always present a discrete change, the change of prediction horizon is a continuous process. A discrete change will occur at the transition of one horizon bin into the next. The selection of bin width for such continuous parameters then is a compromise between sufficient data points to generate an accurate distribution, and the magnitude of the step change of the shape of distribution at the bin transition.

TABLE I. WIDTH, AND CHANGE IN WIDTH, OF THE PDF AT BIN TRANSITION

STA	Horizon (min)				
	60	50	40	30	20
01:30-04:30	21	14 (-7)	12 (-2)	9 (-3)	8 (-1)
04:30-07:30	21	19 (-2)	17 (-1)	16 (-3)	12 (-1)
07:30-10:30	19	16 (-3)	15 (-1)	13 (-2)	12 (-1)
10:30-13:30	17	15 (-2)	13 (-2)	12 (-1)	11 (-1)
13:30-16:30	17	15 (-2)	13 (-2)	12 (-1)	11 (-1)
16:30-19:30	19	17 (-2)	15 (-2)	14 (-1)	12 (-1)
19:30-22:30	16	15 (-1)	13 (-1)	13 (0)	11 (-2)
22:30-01:30	10	10 (0)	11 (1)	13 (2)	14 (1)

value is width of 95% confidence interval

In a proposed display, the main visual feature of uncertainty is the width of a flight's arrival time PDF[21]. To investigate the effect of the transitions, the width of the 95% confidence interval is analysed in each transition. Table I provides the width of the PDF for a number of Scheduled Time of Arrival (STA). It also shows the change at each moment that the predicted time to fly crosses a horizon bin.

At an interval of 10 minutes, the step size can be up to a third of its width (e.g. from 30 to 20 minutes in the first row).

In a visual context, these would cause relatively large visual changes for an otherwise continuous process.

VI. DISCUSSION

A method of predicting arrival time uncertainty using available estimates is demonstrated in this paper. However, the use of FUMs implies that prediction errors included in those messages are inherited by the proposed method. Two of the most important error sources are the influence of the arrival airspace, and the effective resolution of the estimate.

Since the reported landing time was taken as the truth data, any difference between the assumed arrival route and the actual route will result in a prediction error. In predicting, the NMOC assumes that the aircraft follow published approach routes. At AAS, aircraft follow these published routes only during night time. During day time, Air Traffic Controllers (ATCOs) will direct the traffic. The resulting uncertainty in landing time is not meaningful in the context of AMAN as the controllers, amongst others, will use AMAN as an input in directing the traffic. Future concepts may address this inaccuracy by comparing times over particular waypoints.

Since NMOC provides an update only when the actual progress of the aircraft deviates more than five minutes from the prediction, errors below this limit are not taken into account, and are therefore not presented in the data. This problem is partially addressed by the interpolation, thus generating sampled errors of the accurate predictions between the first prediction and the arrival.

A. Other airspaces

Since the data demonstrate that departure uncertainty is the primary cause of uncertainty on the arrival time, both the airline departure accuracy, and the delays at airports generate the larger part of the uncertainty. These two factors account for 20-40% and 5-10% of arrival time deviation respectively, and depend on the airline and the origin airport [17]. This effect is exacerbated through reactionary delay, which is the delay caused by the delay of the inbound aircraft or inbound transfer passengers. These will cause airlines with numerous rotations between two airports, or with large numbers of connecting passengers, to have extra arrival time uncertainty. The arrival time accuracy of the flights are therefore strongly dependent on the city pairs and thus the arrival airports. Therefore, a model estimated for one airport will not necessarily be applicable to another airport, and will require an estimate for each particular case.

B. Other data sources

The FUM was selected since it is currently available and recorded at LVNL. Secondly, only the three most common flight statuses in those messages were used. The NMOC currently provides a more detailed report in the form of the EFD message. This message uses the same data source, and is provided under the same conditions. However, the message provides more detailed routing information, and therefore more possible truth data to determine the prediction error.

Collaborative Decision Making (CDM) provides more details on the progress of the flight toward take-off. Since the data show high uncertainty mainly on departure, inclusion of CDM information might allow to receive more detailed flight states. Unfortunately, the number of origin airports that provide CDM information was too low to provide sufficient data for this research.

C. Bin selection

The parameters of interest, and their respective bin widths were based on the amount of available data points. The current selection of properties was based on an initial analysis, combined with the likelihood that those properties would have an influence. Section III describes a further set of parameters that may be selected. The potential list of properties, as given in Section III provides further options for investigation. For example, it may be worthwhile to distinguish between departure airports by separating flights from the most common airports, and grouping the remainder together. This would enable modelling of the uncertainties due to specifics at those departure airports. Further detailing can come from information available in other predictions, such as CDM messages or EFD.

For simplicity, bin sizes were selected at regular intervals. Section V describes these bin widths and shows that the transition of a bin limit can joined by a relatively large reduction of the width of the PDF. Especially in visual presentation, such instability may be an issue in acceptance by ATCOs. To reduce this problem, the size of bins may well be adjusted based on the other criteria. For example, uncertainty as a function of STAs is likely to be more variable during daytime then during night time, in particular for airports with wave-like traffic densities such as AAS.

VII. CONCLUSION

This paper presents a method for predicting uncertainty in arrival time at longer horizons using readily available data at AAS. By fitting Johnson curves to errors derived from FUMs, a model is generated that is able to more precisely describe the shape of the distribution, in particular skew and kurtosis of the uncertainty. By using the Johnson distribution from tabulated data based on a number of parameters on the received prediction, an error distribution can be calculated fast and is therefore suitable for on-line applications such as displays for ATCOs.

REFERENCES

[1] N. Hasevoets and P. Conroy, "Arrival Manager - implementation guidelines and lessons learned," EUROCONTROL, Brussels, Tech. Rep., 2010.

[2] A. Barff, B. Favennec, P. Conroy, L. Bellesia, J. S. Greenwood, A. Clark, °A. W. Wall, Y. Matsson, S. T¨orner, D. Chouvet, D. Nieuwenhuisen, E. Westerveld, and A. Linner, "SESAR P05.06.04 - D28 – Preliminary OSED Ed. 00.01.01," SESAR Consortium, Tech. Rep., 2012.

[3] J. Bronsvoort, G. McDonald, M. Paglione, C. Garcia-Avello, I. Bayraktutar, and C. M. Young, "Impact of missing longitudinal aircraft intent on descent trajectory prediction," in 30th IEEE/AIAA Digital Avionics Systems Conference, Seattle, WA, Oct. 2011, isbn: 978-1-61284-798-6. doi: 10.1109/DASC.2011.6096062.

[4] SESAR JU, "SESAR Concept of Operations at a Glance ED 02.00.00," SESAR JU, Tech. Rep., 2011.

[5] JPDO, NextGen Integrated Work Plan: A Functional Outline, 2008. [Online]. Available: http://www.jpdo.gov.

[6] J. Bronsvoort, G. McDonald, R. Porteous, and E. Gutt, "Study of Aircraft Derived Temporal Prediction Accuracy using FANS," in Proceedings of the 13th Air Transport Research Society (ATRS) World Conference 2009, Abu Dhabi, UAE, 2009.

[7] E. R. Mueller and G. B. Chatterji, "Analysis of Aircraft Arrival and Departure Delay," in AIAA Aviation Technology, Integration, and Operations Conference (ATIO), Los Angeles, CA, Oct. 2002.

[8] M. Whiteley, "PHARE Advanced Tools Tactical Load Smoother Final Report," Eurocontrol, Brussels, Belgium, Tech. Rep. November, 1999.

[9] D. Nicholls, P. Marti, P. Battino, S. Pozzi, and V. Barsotti, "CARE Innovative Action Presenting Uncertainty to Controllers & Pilots," Eurocontrol, Tech. Rep. December, 2004, pp. 1–53.

[10] T. Mueller, J. Sorensen, and G. Couluris, "Strategic Aircraft Trajectory Prediction Uncertainty and Statistical Sector Traffic Load Modeling," in AIAA Guidance, Navigation, and Control Conference and Exhibit, Reston, VA, Aug. 2002, isbn: 978-1-62410- 108-3. doi: 10.2514/6.2002-4765.

[11] G. Tobaruela, P. Fransen, W. Schuster, W. Y. Ochieng, and A. Majumdar, "Air traffic predictability framework - Development, performance evaluation and application," Journal of Air Transport Management, vol. 39, pp. 48–58, 2014. doi: 10.1016/j.jairtraman.2014.04.001.

[12] R. A. Coppenbarger, "Climb Trajectory Prediction Enhancement Using Airline Flight Planning Information," in AIAA Guidance, Navigation, and Control Conference, San Diego, CA, 1999.

[13] "Traffic Review 2013," Amsterdam Airport Schiphol, Amsterdam, The Netherlands, Tech. Rep., 2013. [Online]. Available: http://trafficreview2013.schipholmagazines.nl/traffic-review-2013.pdf.

[14] G. Sölveling, "Stochastic programming methods for scheduling of airport runway operations under uncertainty," Dissertation, Georgia Institute of Technology, Jun. 2012. [Online]. Available: http://hdl.handle.net/1853/44842.

[15] C. Wanke, S. Mulgund, D. Greenbaum, and L. Song, "Modeling Traffic Prediction Uncertainty for Traffic Management Decision Support," in AIAA Guidance, Navigation, and Control Conference and Exhibit, Reston, VA: American Institute of Aeronautics and Astronautics, Aug. 2004, isbn: 978-1-62410-073-4. doi: 10.2514/6.2004-5230.

[16] H. Koolen and S. Smidt, "Flight Progress Messages," Eurocontrol Network Manager, Brussels, Belgium, Tech. Rep., 2014.

[17] Eurocontrol CODA, "Delays to Air Transport in Europe – Annual 2013," Eurocontrol, Brussels, Belgium, Tech. Rep., 2014.

[18] N. L. Johnson, "Systems of Frequency Curves Generated by Methods of Translation," Biometrika, vol. 36, no. 1/2, pp. 149–176, Jun. 1949, issn: 00063444. doi: 10.2307/2332539.

[19] I. Hill, R. Hill, and R. Holder, "Algorithm AS 99: Fitting Johnson curves by moments," Applied Statistics, vol. 25, no. 2, pp. 180–189, 1976.

[20] D. Jones, "The Johnson Curve Toolbox for Matlab: analysis of non-normal data using the Johnson system of distributions," College of Marine Science, University of South Florida, St. Petersburg, FL, Tech. Rep., 2014. [Online]. Available: http://www.marine.usf.edu/user/djones/.

[21] M. Tielrooij, C. Borst, D. Nieuwenhuisen, and M. Mulder, "Supporting arrival management decisions by visualising uncertainty," in Proceedings of the SESAR Innovation Days, D. Schaefer, Ed., Stockholm, Sweden, 2013, isbn: 978-2-87497-074-0.
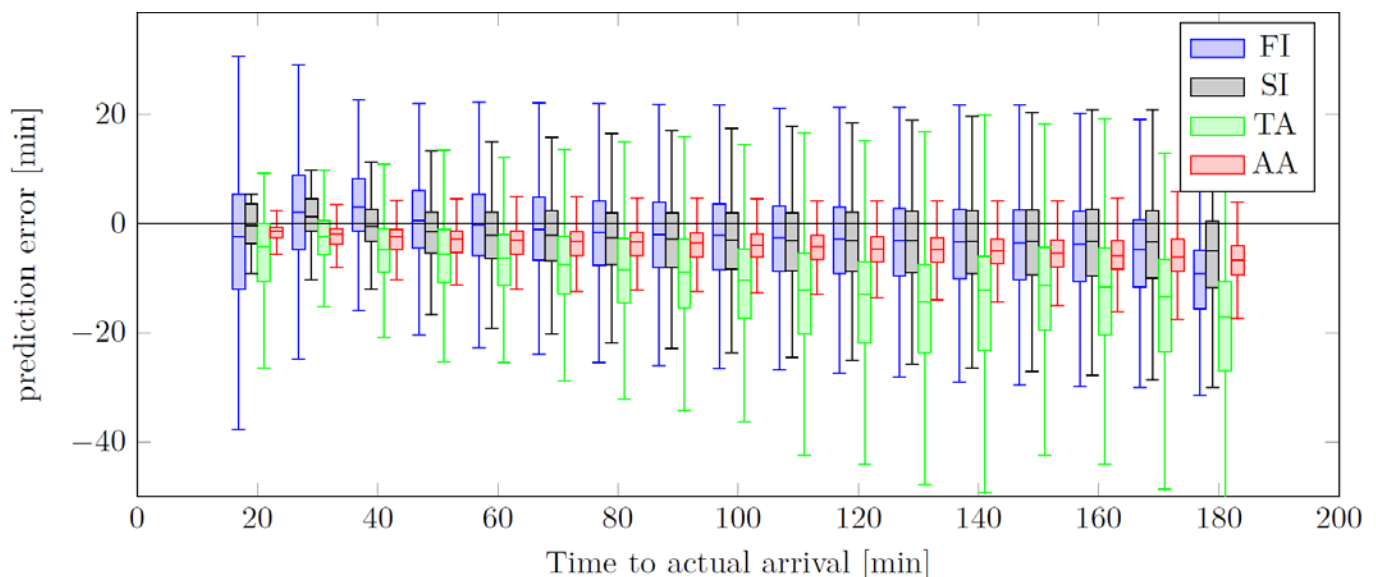
Fig. 4. Accuracy of predictions versus remaining time to fly. Boxplots indicate quartiles, whiskers indicate extreme value within 1.5 IQR from the quartile. Due to the high amount of data points outliers do exist, but have been omitted from the graph for clarity.
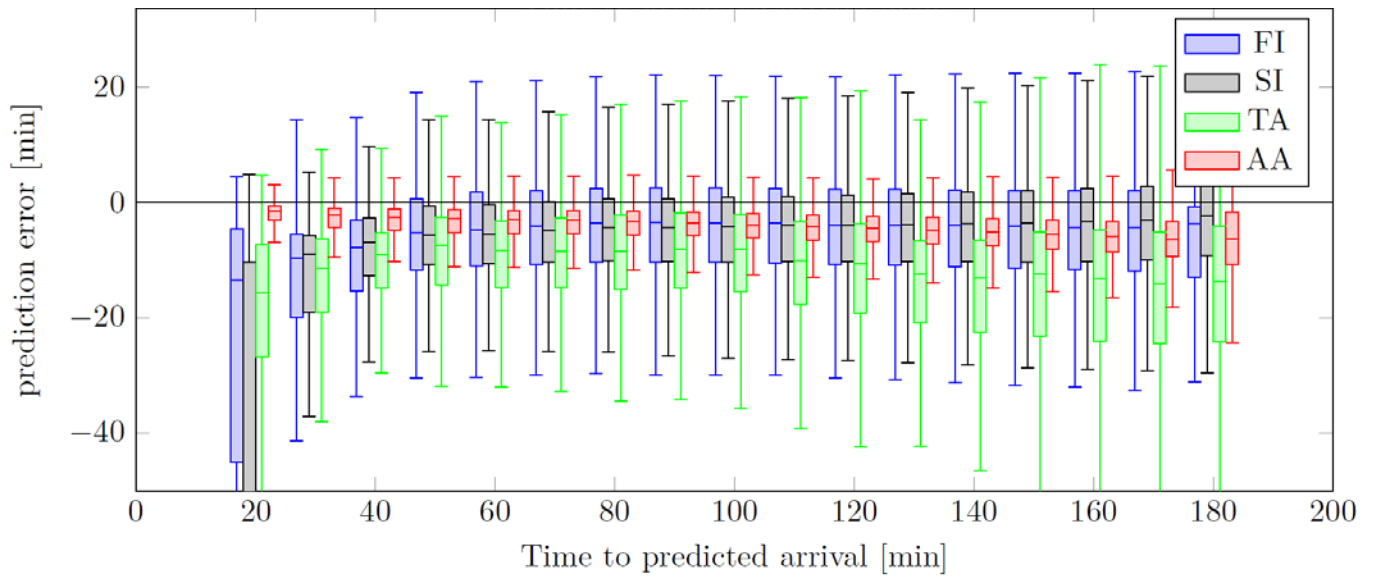
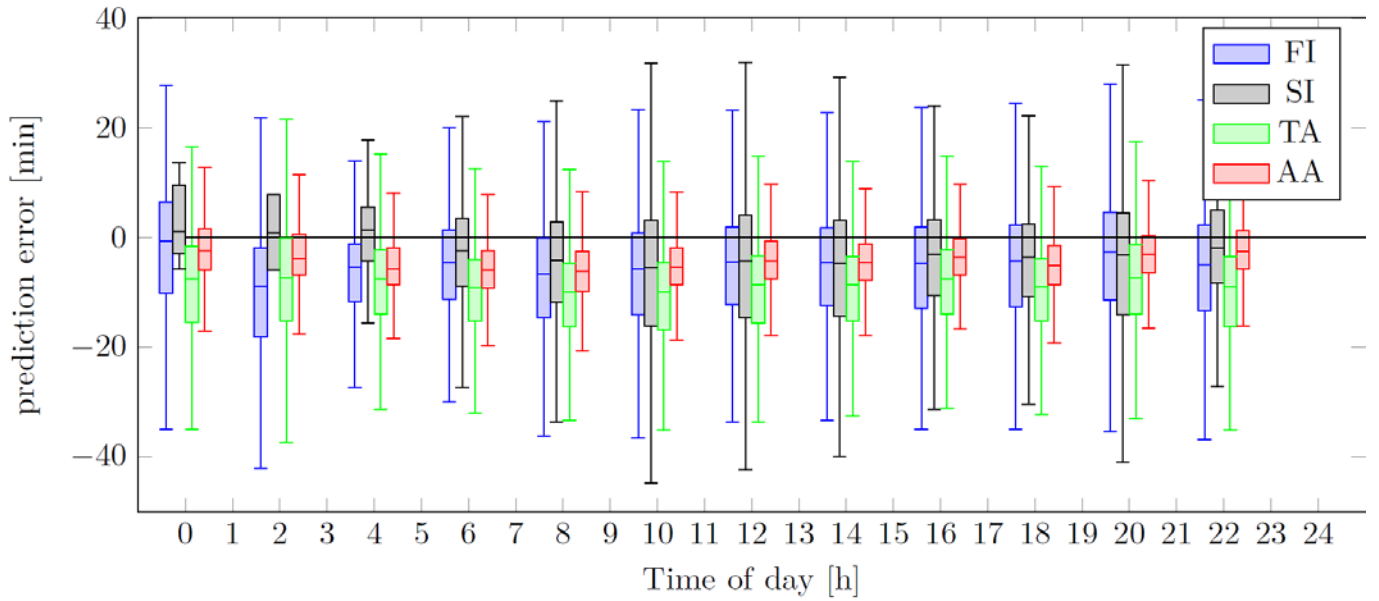Fig. 5. Accuracy of predictions versus predicted time to fly.

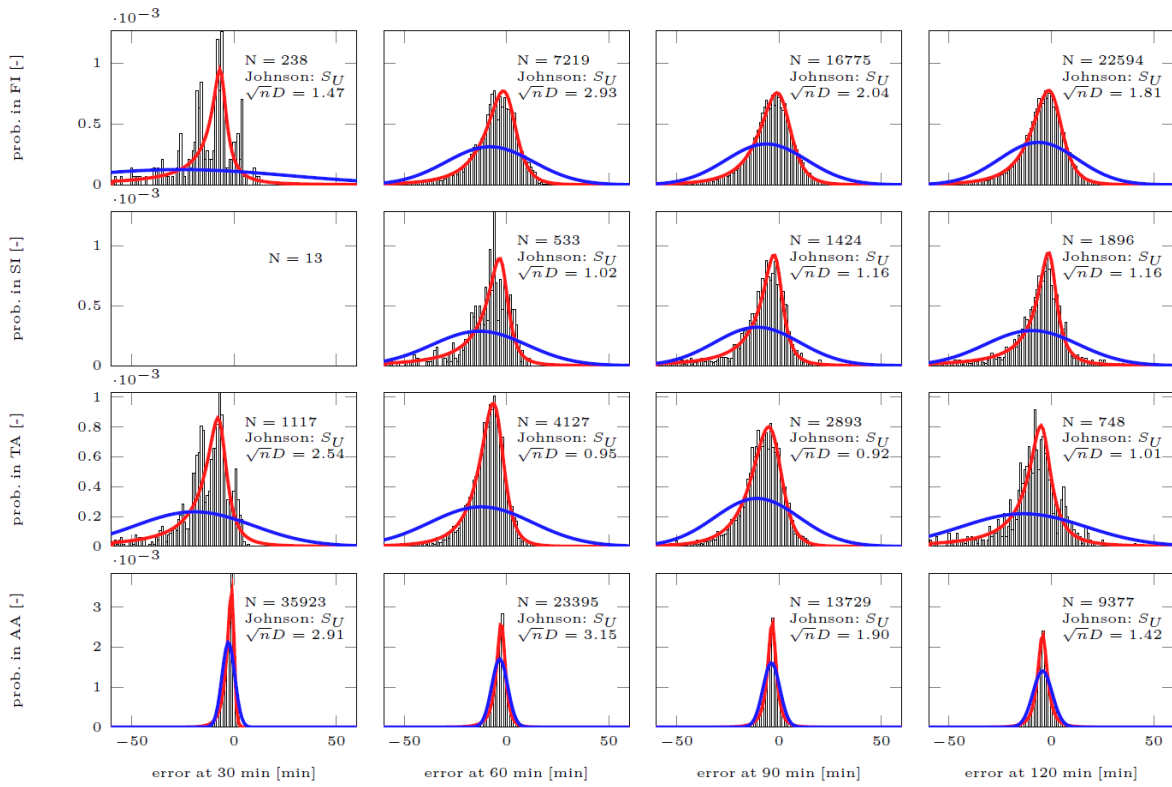Fig. 6. Accuracy of predictions versus the scheduled arrival time.

Fig. 7. Distribution of predicted error (in minutes) as function of flight status and predicted time to arrival for flights scheduled to arrive between 10:30 and 13:30. The annotation show the number of samples, the type of Johnson curve applied (SU for unbounded, SB for bounded) and the Kolmogorov-Smirnov test statistic scaled by the number of samples.
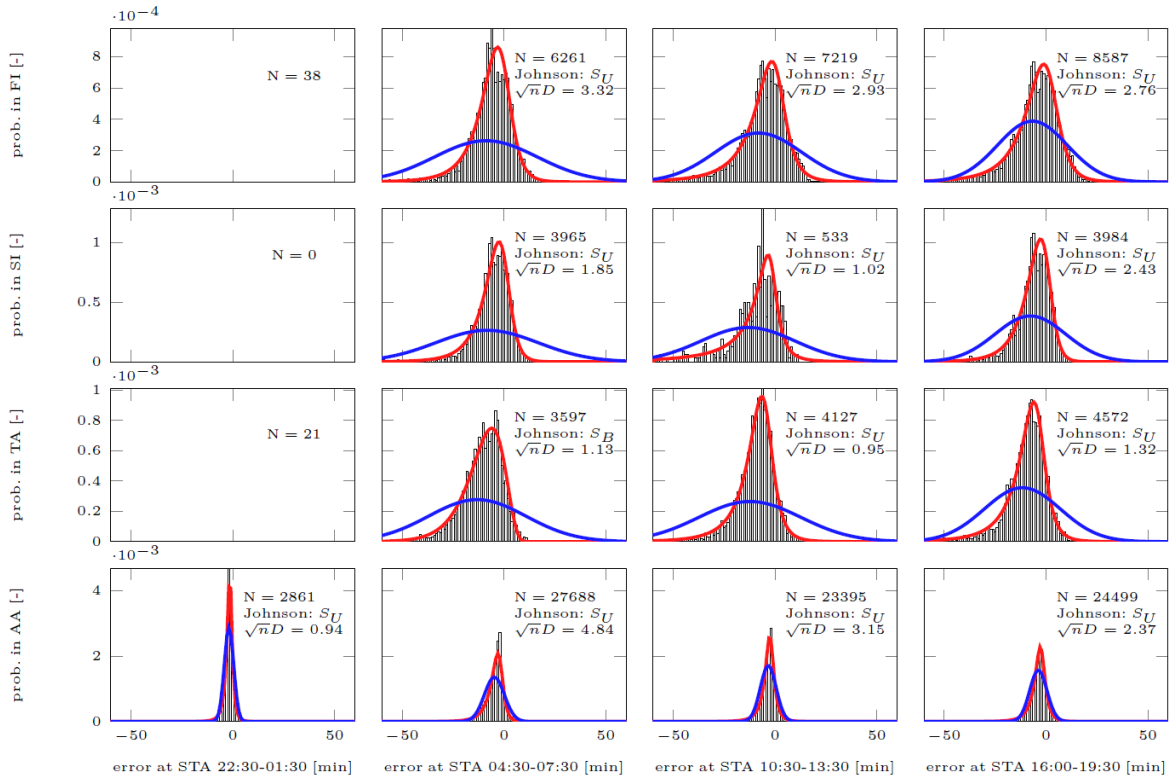

Fig. 8. Distribution of predicted error as function of flight status and planned time of arrival at a prediction horizon of 60 minutes.