A machine learning approach for the prediction of top of descent

Tan, Benjamin Zhi Yong; Alam, Sameer; Ma, Chunyao

2021

Tan, B. Z. Y., Alam, S. & Ma, C. (2021). A machine learning approach for the prediction of top of descent. 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC).
<https://dx.doi.org/10.1109/DASC52595.2021.9594470>

<https://hdl.handle.net/10356/151957>

<https://doi.org/10.1109/DASC52595.2021.9594470>

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/DASC52595.2021.9594470>.

Downloaded on 26 Apr 2023 13:35:53 SGT

A Machine Learning Approach for the Prediction of Top of Descent

Tan Zhi Yong Benjamin, Sameer Alam, Chun Yao Ma

Air Traffic Management Research Institute,

Nanyang Technological University, Singapore

benjamin.tanzy@ntu.edu.sg, sameeralam@ntu.edu.sg, M180146@ntu.edu.sg

Abstract—Under the light of increasing air traffic congestion and reduced environmental impacts, the growing need for tailored approaches has arisen. To coordinate tailored approaches, accurate predictions of descent trajectories must be available. The ability of ground based automation for such endeavours however must be improved. Specifically in the ability to predict where descents begin, also referred to as the Top of descent (TOD). This paper attempts to provide a machine learning approach to TOD location prediction, considering the operational capabilities of modern-day air traffic management by working within the confines of parameters available to controllers and considerations of stakeholders. To identify the common TOD clusters and the resultant average TOD locations representing the controllers macroscopic perspective, an unsupervised learning approach is utilised. Specifically, an ensemble method of DBSCAN-K-means is used. This was compared to the more microscopic pilot controllers perspective and generic decision-making formulas to determine the proximity to expected TOD. Of which the microscopic option has proven to be more reliable in representing the individual Flight Management System (FMS) recommendations or pilots' decisions on aircraft. In the prediction of TOD location, operationally available factors such as proximity to expected TOD, current flight level, approach heading, ground speed, fleet and aircraft type are used to train a decision tree model. A Singapore flight information region (FIR) case study is conducted to illustrate the methodology. Results of the methodology employed on the inbound A380 fleet consisting a total of 79 flights, yielded a maximum prediction range of 31NM range from actual TOD occurrence. This paper has demonstrates the possibility of using a machine learning approach to predict TOD of a flight considering both prior cruise characteristics and operational considerations of pilots.

Index Terms—Machine learning, Top of descent, Singapore, TMA

I. INTRODUCTION

Within the congested airspace environment, ever-increasing projected traffic and their related environmental impacts continually push the transition from conventional arrival procedures to more direct approaches [1], where continuous descent approaches (CDAs) with low to minimal thrust is best desired. Compared with conventional staggered descent approach which involves usage of thrust to maintain speed to hold altitude, CDAs advocates a constant rate of descent approach direct for landing. Such approaches are exemplified at airports such as London Heathrow airport, where the lack of thrust usage allows for better noise abatement, reduced fuel burn and overall lowered operational cost per flight [2]. While CDAs are the ideal, this is often affected by several

factors limiting their use; such as weather mitigation, air traffic deconfliction as well as aerodrome management [3]. This has given rise to the concept for tailored approaches, where inbound flight trajectories are preemptively managed during and prior to the descent phase to collectively ensure reduced deconflictions, delays and increase efficiency by remaining close to CDA operations. One emerging solution to optimising arrival operations is the use of Extended Arrival Manager(E-AMAN) systems which increases the arrival management horizon up to 180-200 nautical miles away from airport [4] enabling better flow management and sequencing.

Enabling such endeavors however, require accurate ground-based automation that can predict the descent trajectories of aircraft [5]. As CDA is heavily reliant on when the descent is commenced, the prediction of TOD is the first crucial point to be investigated. While good predictive factors have been investigated such as aircraft drag, lift to weight ratio and prevailing winds [6], [7], acquisition and fidelity requirements of data may not be readily available for ground-based operations. Moreover, given the stochastic nature of aviation, an apparent randomness in TOD locations can also be expected. Under the light of data being subject to noise, lower recording rates and apparent randomness, the use of machine learning can provide an alternate solution to prediction of TOD location due to its ability to remove or minimise the impact of noise in data, understand unknown patterns between inputs and make predictions even with limited data or parameters. [8]. As such, this paper attempts to provide a machine learning approach to the prediction of current Top of Descent (TOD) location based on prior cruise characteristics and operational considerations, encapsulating the prior investigated factors such as aircraft type as latent contributors.

For this research, an attempt to exploit the use of unsupervised machine learning on historical data is done via an ensemble method of DBSCAN-K-means [9]. This yielded the common TOD clusters and the resultant average expected TOD locations based on the approach direction, fleet type. These macroscopic average expected TOD positions were then compared to the more microscopic pilot's operational perspective using commonly referenced formulas [10] to determine a suitable prediction factor namely, the proximity to expected

TOD. Next, in the prediction of TOD location a decision tree model is trained and validated using the defined operational parameters alone. Regarding training and prediction of the supervised learning model, the parameters for proximity to expected TOD, current flight level, approach heading (associated flow direction), ground speed, fleet/aircraft type and landed runway are used. To illustrate the proposed methodology, an experimental study was conducted for Singapore flight information region (FIR) because of its perceived complexity in estimating TOD location stemming from its diversity in fleet and airline operators and multiple possible approach headings. Results of the methodology employed on the inbound A380 fleet in Singapore FIR yielded a maximum prediction error range of +31NM (equivalent to 4 minutes of flight time) when estimating TOD location. Through this research, it has been shown that the use a machine learning approach to predict TOD through prior cruise characteristics and operational considerations alone is feasible. The rest of the paper is organised as follows. Section 2 details the background of existing works. In Section 3 the methodology employed to predict TOD is presented. Section 4 illustrates the experimental case study results of the methodology as well as the evaluation of results. Lastly, future recommendations are made before the paper is concluded in section 5.

II. BACKGROUND- FACTORS AFFECTING TOD

This section presents in brief the existing works and considerable factors regarding TOD estimation for use in this research.

1) *Aircraft type affecting TOD*: In the research done by Craig M. Johnson, [6] it was highlighted that one area of concern in a flights trajectory prediction stems from an aircraft's TOD. To paraphrase, misleading TOD information given by erroneous automation can rapidly develop into later loss of separation scenarios. With the help of automation, the on-board FMS calculates an expected TOD location for an idle or near idle thrust descent. This location is determined according to aircraft performance, the environmental climate (e.g. weather and winds), airspace constraints and preferred operator settings such as cost index which affects the thrust usage and descent speeds allowed. With the intent of increasing safety and efficiency in National Airspace System (NAS), the research sought out to identify and understand which other factors significantly affect variance in TOD. From the given results in [6] surrounding the Boeing 737 airframe, it was concluded that an older series of aircraft tended to descend further from their fix as compared to newer series aircraft. Moreover, the presence of winglets tended to cause aircraft to descend 2 to 3 nautical miles further from fix and proved to be a significant factor for TOD differences. Lastly, engine thrust factor was found to not have a significant factor for TOD differences. This proved differences in aircraft structure and type are factors to be considered in flight trajectory prediction performance.

2) *Aircraft performance affecting TOD*: From Laurel Stell's research [7], an investigation to find the factors affecting

TOD location was performed to facilitate development of the Efficient Descent Advisor (EDA) ground predictor tool by developing several polynomial approximations of the distance between TOD and its meter fix. Results yielded that several parameters were significant in prediction of TOD location. Specifically, descent speed, aircraft weight and wind speed. Other fixed factors were also used such as cruise altitude, aircraft type, altitude and speed required at the given metering waypoint(fix). From application, the author notes the need for accurate value requirements in order to meet the EDA maximum allowed prediction error. With perfected inputs, attempted prediction of simulated B737 TOD, resulted in a 2-6NM prediction range from FMS computed TOD falling in range of the maximum allowed absolute error goal of 5NM.

III. METHODOLOGY

In this section the proposed methodology for a machine learning method in predicting TOD location. Figure1 demonstrates the overview of the proposed method.

Fig. 1. Methodology diagram

As illustrated in Figure1, input historical flight data is pre-filtered and processed for only arrival flights (yellow). After which the data relevant to the input layer of the supervised learning model is collated (green), this layer includes the cruise parameters prior to TOD location and a derived parameter describing the aircraft proximity to its expected TOD location calculated based on either the historical average (macroscopic) or from the individual (microscopic) perspective, mimicking the ATC's or pilot's respective knowledge and decision making. Both methods to determine Proximity to Expected TOD Parameter ($TOD_{DIFF}(macro)$ and $TOD_{DIFF}(Micro)$) are further elaborated in subsections B and C below. The input layer is then fed into a supervised machine learning model for training, followed up by prediction of TOD locations for a particular flight (blue). The methodology is finally concluded with a result validation process which

determines the accuracy of the trained model. The subsections below describe in further detail the specifics of each process undertaken.

A. Overview of Data input

1) *Utilised Data Source*: For this research the use of Automatic Dependent Surveillance–Broadcast(ADS-B) data is proposed. As the current replacement for Secondary surveillance data [11], ADS-B provides one of the best possible data sources when working within the limitations and confines of ground based operations. While lessened in detail as compared with higher fidelity data sets such as the aircraft accessed Quick access recorder (QAR) data. Using the individual aircraft characteristics and 4D trajectory, ADS-B can easily be used to determine various other parameters such as phase of flight, landed runway, the departed flight level prior to descent and distance from arrival airport [12]. The primary parameters utilised within the proposed dataset is listed in Table I with their respective definition. In addition, the use of other parameters such as flight Identification code and aircraft registration number is recommended for ease in distinguishing between flights.

TABLE I
ADS-B REQUIRED PARAMETERS

Required parameter	Units	Definition
Latitude	degrees, minutes and seconds (DMS)	Horizontal positioning parallel to equator, GPS coordinates
Longitude	degrees, minutes and seconds (DMS)	Vertical positioning perpendicular to equator, GPS coordinates
Heading	Degrees (°)	Aircraft magnetic heading
calibrated altitude	Feet (ft)	Indicated altitude corrected for instrument error
Ground Speed	Knots (Kt)	True airspeed accounted for relative wind
Aircraft type	NIL	Model type of the aircraft
ICAO flight label	NIL	assigned flight call sign
UTC time	UTC	10 digit coordinated Universal time

2) *Data preprocessing*: As TOD locations are only applicable to arriving flights, the intended ADS-B dataset is first filtered to separate individual flight trajectories based on the intent within the FIR, where departing and en-route flights are removed. In addition, arriving flights with ambiguous intent or incomplete trajectories are also removed due the unknown destination airport or trajectory details. the addition of these trajectories may lead to the inevitable inclusion of arrival flights of nearby airports reducing accuracy in the supervised learning model's training and prediction.

3) *Parameter generation*: To develop the required input parameters, algorithms were developed to determine the altitude differences in flight, distance from threshold, estimate the flight level, TOD location and landing runway with regards to each individual flight trajectory.

To elaborate, the altitude difference parameter is a simplistic vertical situation parameter for a particular flight. Defined as the difference between one recorded altitude data point and the succeeding altitude data point one can easily determine if the flight is on descent (positive value) in a climb (negative value) or in level flight (zero) similar to the on-board vertical speed indicator (VSI). While the vertical speed can be determined by dividing the difference across the time taken, the low recording rates of ADS-B data often lead to inaccurate values which stands in opposition to the findings of Stell. L [7] who states the need for high accuracy requirements in the vertical speed parameter for use in accurately predicting TOD location. In addition, using a similar manner, the inter-point distances and distance from landed runway threshold (distance to threshold) can be calculated as well. Another parameter to be derived is the estimated flight level; When used in tandem with the prior found vertical situation, the respective TOD location for each flight can also be found.

Algorithm 1: Identification of TOD location

Data: Individual historic flight trajectory

Input: Flight parameters: Length of data (len),
Vertical situation (VS), Altitude(A), Transition
altitude (TA), minimum duration for descent
(n)

Output: True/False evaluation of TOD
location/occurrence

```

initialization;
for  $i = 0$  to  $len$  do
    if  $VS[i - 1] = 0$  then
        if  $VS[i, i + 1...n] < 0$  and  $A[i + n] < TA$  then
            TOD[i] = True;
        else
            TOD[i] = False;
        end
    else
        TOD[i] = False;
    end
end

```

While TOD location can be found via the vertical situation parameter alone by detecting the transition from level flight to a consistent descent flight. Complexity arises when historical trajectories include a significant flight level changes (descent) prior to TOD. Both TOD and change in flight level can be described as a prior level flight and a period of descent after the decision location. Yet, one observable difference can be used to accurately define TOD locations for a flight. Where flight level changes have a descent component, the end of this descent remains within the en-route airspace specifically. TOD occurrence however ends the descent within the terminal airspace (below transition altitude). Also observable, the duration of descent is often longer as compared to flight level changes. In determining the actual TOD location for each flight trajectory an algorithm is used. This process is illustrated through a

Fig. 2. Creation of runway landed parameter example

generic algorithm presented in algorithm 1. To elaborate, TOD location is determined by first detecting the initiation locations for descents from level flight, the algorithm then also detects for a significantly prolonged period of descent, lastly the algorithm checks for the termination of descent in terminal airspace. Should the above mentioned conditions be met, TOD at that initiation location is said to be True. Lastly, a parameter is developed to further filter the dataset according to the respective landed runway. This is because the conducted approach and descent often vary depending on the runway in use. Figure 2 is a pictorial example of how landed runway parameter is determined. Using a point-in-polygon method, geographical gates are first set to detect the aircraft's proximity to the arrival airport (first gate), after which at the altitude where aircraft are commonly expected to intercept the Instrument Landing System (ILS); estimated to be between 4000ft to 2000ft, the heading of aircraft is finally used to determine the runway landed at. Where several runways exist, the point-in-polygon method is again employed (second gate) to determine the specific runway a flight landed on. While this parameter is taken from historical data for training, in prediction of TOD during real time operations, the runway in use for arrivals should be used as the alternative as the actual runway aircraft will land on has yet to be determined.

4) *Proximity to Expected TOD*: In an effort include the stakeholder's perspective of Airlines and Air Navigation Service Providers (ANSPs) as a input parameter, the use of unsupervised machine learning to determine if there exists a discernible link between the aircraft model type, approach heading and TOD location is exploited. This is done by determining the expected TOD location of each flight and their proximity to said location. Adopting the controller's perspective of where they would expect TOD of flights to occur. A clustering algorithm to determine the average expected TOD location is used. As a myriad of factors can change where TOD locations occur, the earlier described macroscopic perspective

is also compared to the microscopic viewpoint of the pilots training and decision making tools available. As exemplified by Adriana and Uemura [13] the formulas employed by the individual Flight management system (FMS) of each aircraft type should be primarily referenced. However, because they remain proprietary their use is often hampered and limited. Moreover, as the FMS formulas require the accurate thrust-drag ratio, desired rate of descent and exact weight of aircraft to determine TOD, ground based operations may often find itself lacking the data required. As such, alternative commonly referenced formulas must be employed in order to imitate the FMS in a generic manner. In the proposed methodology utilisation of the CDA assumption formula commonly referenced by pilots [14] is recommended. Both methods to determine Proximity to Expected TOD Parameter ($TOD_{DIFF}(macro)$ and $TOD_{DIFF}(Micro)$) are elaborated in the following subsections below.

B. Macroscopic Expected TOD via ensemble clustering

As described, it is attempt to include the air traffic controller's empirical knowledge of when aircraft would begin its descent. In simpler terms, while not quantitatively expressed it is not uncommon for an experienced controller to use prior experiences and commonly seen behaviours to estimate where TOD would occur. To best quantify this "sensed" behaviour, an ensemble clustering model is used to give an averaged location of TOD. For this research a DBSCAN-K-Means model is employed to experiment its suitability. As seen in Figure 3, the process is described as follows:

Fig. 3. Process flow diagram - $TOD_{DIFF}(macro)$

1) *Flight trajectory and cluster visualisation*: First, the collective flight trajectories and found TOD locations are first visualised on a 3D (latitude, longitude, altitude) and 2D axis (latitude, longitude), (distance to threshold, altitude). From the resultant visualisation model, the flow structure of the airspace is presented. Analysing the flows approach headings, initial observations and operational knowledge can be employed to

determine how many and where TOD clusters exist along an approach direction.

2) *DBSCAN*: While visually acknowledged TOD clusters may be easily seen for a controlled dataset, different scenarios and conditions encapsulated by an alternate datasets may cause significant variations in TOD locations resulting in less distinguishable common clusters. This can lead to confusion to where exactly clusters exist and how many there are. To overcome this hurdle, the first half of the ensemble clustering model is utilised to distinguish the common clusters of existing TOD locations and exclude outliers. Identifying clusters based on the density of points, DBSCAN first considers the minimum proximity between 2 points defined by the epsilon value (*eps*) to be included in a proposed cluster. Next, the minimum sample value (*min_{sample}*) is then used to define how many samples are needed in a proposed cluster in order to actual be classified as a cluster. This model not only quantifies the previously ambiguous number of clusters, it also ensures the clusters found are indeed the areas where TOD is commonly conducted, eliminating the outliers. By varying *eps* and *min_{sample}* the evaluated clusters can be adjusted to determine the extent of outlier exclusion. An example is shown in Figure 4 showing how varying *eps* and *min_{sample}* affects cluster size and inclusion of points.

Fig. 4. Sample results of varying *eps* and *min_{sample}*

3) *K-Means*: Common clusters of TOD found by DBSCAN can be unsymmetrical, leading to difficulty in mathematically calculating the average TOD location of each respective cluster. Through the second half of the ensemble clustering model, the K-Means models simplifies the process by determining the centroid of the cluster. Before using K-Means, two things must be defined, number of clusters(*k*) and positioning of *Kvalues* (initial starting points of each cluster's centroids). To determine the centroid location, the K-Means algorithm then references the *Kvalues*, linking TOD data points with its nearest centroid, identifying a starting cluster. By taking a new mean of all identified points in each cluster, re-clustering is performed. when centroids are stabilised, iterations stops and actual cluster centroids are identified. Although *Kvalues* can

be randomly generated, user-defined placement of these points based of previous observational links (such as which common flow or approach heading a specific cluster encapsulates) reduces the possibility that a cluster is wrongly identified, ensuring the final iteration of *Kvalues* is the exact centroid of an actual cluster. Regarding *k*, the found results from DBSCAN is used as an alternative to the optimal *k* elbow method.

4) *TOD_{DIFF}(macro)* parameter: From the ensemble clustering model described, the macroscopic expected TOD locations for each cluster is found. Referencing the approach heading prior to TOD, a particular flight is linked to its most relevant cluster and expected TOD location (centroid). For each flight's trajectory, the *TOD_{DIFF}(macro)* parameter is given by an algorithm used to calculate the difference in distance between a) the distance to runway threshold value at expected TOD location and b) distance to runway threshold value at actual TOD respective to the fleet or aircraft type and approach heading.

C. Microscopic Expected TOD via Pilot decision mimicry

As the alternative to the perspective of controllers, the Microscopic viewpoint of the pilots is also considered.

1) *CDA assumption formula*: Explained earlier, while actual FMS formulations may be the better choice to estimate an aircraft's expected TOD location, its hampered use gives way to an alternative method that provides a generic but simplistic replacement. From Pilots, it was understood that they utilise the CDA assumption method for fast evaluation in decision making [15]. One such example is given by the formula:

$$TOD \text{ lateral distance} = \frac{3 (\text{Departing flight level})}{10} \quad (1)$$

This generic rule first references the departing flight level, prior to the descent phase. For historical flights this is found by referencing the last recorded flight level prior to TOD. In real time calculations, the intended or current flight level should be used. This value is then multiplied by 3 then divided by 10 resulting in the estimated lateral distance in nautical mile(NM) from landing runway threshold required to commence descent in order to match the 3-degree glideslope expected for a continuous descent operation. This approach do not consider level restrictions at the various waypoints/STAR fixes in TOD computations. The CDA computation method determines the distance from landing runway threshold using a "straight-in" approach, as opposed to the altitude and speed restrictions of Standard terminal approach routes (STARs) metering fix. This is done given the variability of flight trajectories in the TMA. While the CDA computation formula can be modified to take the metering fix restrictions into regard, associating a flight trajectory to its relevant STAR route and metering fix can be highly varied due to vectoring, track shortening, etc. The benefit of this approach is the ability to still produce an individual expected TOD location for each flight based on its prior cruise parameters without the exclusion of outliers,

unlike the prior method, where flights are assumed to descend in a relative common location.

2) $TOD_{DIFF}(micro)$ parameter: Similar to it's sister parameter, the $TOD_{DIFF}(micro)$ parameter is given by the difference between the distance to runway threshold value at actual TOD and calculated expected TOD lateral distance from runway. The output of the clustering model however is also used, where the most relevant cluster is identified as an input variable for the supervised model.

D. Supervised machine learning via decision tree

Exploring the use of a machine learning in predicting TOD, supervised learning model is employed. For this research a decision tree model is used in order to produce an explainable algorithm and reduce time required for training. Utilising a continuous variable decision tree in SCIPY software, a flow chart-like process occurs. This supervised machine learning process works on the concept of true or false across several levels to predict an output based on prior data input. To decide which feature is placed at the root node or the subsequent internal nodes, Gini Index is used [16], given by the formula:

$$Gini = 1 - \sum_{i=1}^n (P_i)^2 \quad (2)$$

Where P_i is the probability of an object being classified to a particular class

The Gini Index value is found between 0 and 1, where 0 is representative of all elements belonging to a specified class or a single class exists alone and 1 representative of a random distribution of elements across various classes. A Gini Value between 0 and 1 such as 0.5 symbolises an equal distribution of elements across classes. In building a decision tree, the input variable with the lowest Gini index is used as the root node.

TABLE II
INPUT LAYER VARIABLES

No.	Input variables:
1.	Proximity to Expected TOD (Macro or Micro)
2.	Associated cluster (corresponding approach heading)
3.	Flight level prior to TOD
4.	Aircraft type
5.	Runway used
6.	Ground speed

As seen in table II, to train the decision tree, the following input variable are used. By pre-filtering flight data by their fixed values (associated cluster, runway used and aircraft type) data can be more easily managed and results in a specific and explainable decision tree; For actual prediction of TOD in real time, the current "runway in use" for arrivals is referenced in place of the historically recorded "runway used" parameter.

In preparation of data for supervised learning, input parameters co-relating to actual TOD locations are collated. In addition, sample flights are added to collated TOD input parameters, feeding positive and negative examples to the

algorithm. Because data is filtered according to approach heading and fleet type, During training, it is important for the negative examples used to be of the same approach heading and fleet type as well. Finally, 10 percent test flight trajectory data is set aside for use in model validation with 90 percent used for training of model. In this research Model is assessed based on its ability to accurately predict the TOD of a flight given by the range of it error. For this study, a prediction horizon of 250NM from the arrival airport is used, ensuring the prediction algorithm is trigger in the cruise flight phases of arriving aircraft prior to TOD occurrence.

IV. CASE STUDY

A. Background on Singapore FIR (WSSS)

As illustrated by Figure 5, the complexity in estimating TOD is given from the multiple possible STARs, which represent the many approach directions arrival aircraft can come from (black lines). Paired with its service of a significant number of airlines and respective varied fleets [17] uncertainty in TOD prediction is further intensified. This causes perceived difficulties in predicting TOD and thus hinders growth towards tailored approaches and use of CDA. Due to this, Singapore FIR is best suited to illustrate the proposed machine learning approach.

Fig. 5. WSSS FIR arrival(STAR) flow structure

B. Data input and preprocessing

For this study, 1 month of ADS-B data was obtained via the flight data and tracking website Flight Aware. Contained within the given dataset was 18768 recorded flight trajectories within Singapore FIR, differing in flight phases.

C. Data preprocessing

From the filtering process, en-route flights were noted to contain no more than 30 data points or lasting a maximum record time of 10 minutes per trajectory. The algorithm used filtered out these flights out by referencing the length and duration of the trajectory. Departure trajectories were then filtered out by detecting if aircraft were on a trending

descent at altitudes below 1000 feet, flights that were instead climbing were rejected. Left with arrival trajectories, the algorithm developed yielded a final 7240 arrival flight trajectories. An additional truncating process was also used to only include data relevant to the prediction horizon consisting of TOD, descent, landing and later Cruise phase (50 data points outside of Singapore FIR, estimated to be 250 NM from ICAO:WSSS). After the filtering process, parameter generation was performed on the remaining arrival flights to define altitude differences in flight, distances from threshold, estimated flight levels, TOD location and landing runway with regards to each individual flight trajectory.

D. Flight path visualisation

Using 3D axis plot, the latitude, longitude and altitude were plotted to visualise the trajectories filtering according to the 4 possible active runways: 20C, 02L, 02C and 20R. Respective visualisation plots are presented in Figure 6 with cyan dots representing the runway thresholds.

Fig. 6. 3D glideslope visualisation for runway 02C, 20L and 20C

During this visualisation process, it was noted the obtained dataset did not encompass approaches for 20R and had a small number of arrival flights for 20C. The 2D axis plots were also plotted to depict the collective lateral trajectories profile (distance to threshold, altitude) and horizontal trajectories profile (latitude, Longitude) seen in Figure 7 and 8. From the diagram, all trajectories (blue lines) were first plotted and corresponding TOD locations (red filled dots) were overlaid, with cyan dots representing the runways aircraft landed on.

Fig. 7. Collective 2D horizontal profile (latitude, Longitude)

Fig. 8. Collective 2D lateral profile (distance to threshold, altitude)

From the visualisation process, 3 observations were made. First, it was recognised that pilots exited cruise and entered descent flight phase in relatively close locations, of which an initial estimation of 6 possible TOD clusters possibly exists among all arriving flights shown encircled by red in Figure 7. Secondly from the lateral profile, the conduct of TOD was shown to be within an estimated range between 90NM to 140NM (red lines) with the exception of 25 flights who had abnormal recordings, a higher density in TOD occurrence nearer to the lower bound was also noted. Lastly, inbound flights were observed to initiate their descent according to the CDA approach, only transiting to a staggered approach after intercepting the Initial Approach Fix (IAF) of STAR charts.

E. Macroscopic Expected TOD via ensemble clustering

In determining $TOD_{DIFF}(macro)$ parameter the dataset was further filtered according to aircraft type. For the purpose of best exemplifying the given methodology, the rest of the case study focuses on the A380 fleet alone. From the fleet separated dataset, A DBSCAN model was used to identify the

common TOD clusters. Through an experimental process, an eps of 0.3 and min_{sample} of 5 was found to best balance cluster definition and exclusion of outliers. The visualised sample results for the A380 fleet is presented in Figure 9. From the results, 5 common clusters were found (red, cyan, yellow, green and orange). These clusters were noted to be somewhat similar to the earlier initial observations made in Figure 7 with the exception of one, where no recorded A380 trajectories were available. A confusion matrix is shown in Figure 10 to validate the similar accuracy of clustering via DBSCAN in comparison to the visual assignment of clusters.

Fig. 9. Results of DBSCAN - A380

Confusion Matrix - A380							
Visually assigned label	C1	C2	C3	C4	C5	C6	Outlier
	79	0	0	0	0	0	2
	0	6	0	0	0	0	2
	0	0	32	0	0	0	3
	0	0	0	7	0	0	0
	0	0	0	0	29	0	0
	0	0	0	0	0	0	0
Outlier	0	0	0	0	0	0	26
Predicted Label							

Fig. 10. Confusion Matrix (DBSCAN, visual assignment) - A380

Excluding the outliers, the Common TOD clusters were ingested into a K-Means model to determine the respective centroids. Using the evaluated k of 5, the user defined K values was set within the defined common TOD locations at the estimated centre. Seen in Figure 11, the final iteration

of the each cluster for the A380 fleet is shown, with the respective centroids or expected TOD location highlighted in black. The clusters were also labelled in numerical order for ease in later classification as well. Lastly, $TOD_{DIFF}(macro)$ parameter is created for each trajectory as described in the given methodology.

Fig. 11. Results of DBSCAN-K-Means - A380

F. Microscopic Expected TOD via ensemble clustering

As per the proposed methodology, expected TOD location and consequently $TOD_{DIFF}(micro)$ parameter was also created. In an effort to present the validity of using the formula as a generic alternate for FMS, sample calculations for the highest (FL410) and lowest (FL290) recorded departed flight levels are presented below.

$$TOD \text{ lateral distance (highest)} = \frac{3(410)}{10} = 123 \text{ NM} \quad (3)$$

$$TOD \text{ lateral distance (lowest)} = \frac{3(290)}{10} = 87 \text{ NM} \quad (4)$$

Referencing earlier observations made, it was noted that descents were somewhat similar to CDA approaches with TOD occurrence typically found to be within a range of 90 NM to 140NM. In comparisons with the presented sample calculations, the calculated expected TOD range seemingly fall within the observed range showcasing the ability of the formula to estimate expected TOD location within the vicinity of actual TOD occurrences.

G. TOD_{DIFF} parameter selection

In developing $TOD_{DIFF}(macro)$ and $TOD_{DIFF}(micro)$ parameters, it was observed that using CDA assumption formula yielded a more consistent difference between expected

TOD location and actual TOD occurrence as compared to the macroscopic clustering method. Although it has been found that common clusters of TOD does exist for specific common approach headings and fleet or aircraft types. Using the rigid centroid(mean) of a cluster to determine $TOD_{DIFF}(macro)$ ignores the many contextual considerations that affect the variance in a cluster leading to inconsistent values. Varying possibilities such as weather influence, airline operation procedures and instructional commands from air traffic controllers can contribute to high variance within a cluster as seen in blue and red cluster of Figure 11; regarding such clusters, accuracy of TOD prediction is foreseen to be low and unreliable. While use of the $TOD_{DIFF}(macro)$ is still applicable for some clusters, with lacking contextual consideration input, use of $TOD_{DIFF}(macro)$ is limited. As such the use of $TOD_{DIFF}(micro)$ parameter was opted for the training of decision tree due to its generic formula, consistency and individualised aspect.

H. Decision Tree application

In training of the decision tree model, the dataset was first pre-filtered according to its fleet type and most relevant cluster via its approach heading. While it was stated that the runway parameter should be considered, dividing the already limited dataset would leave insufficient data for training. It was opted to neglect this parameter for the conduct of this study. After filtering process, input data was created by collating data points co-related to actual TOD locations. To provide negative examples for training as well, a sample quantity of non TOD data points (Flight locations where TOD did not occur) was also included in the collated dataset. The presented results in this study is taken from the largest A380 fleet cluster available, cluster 1. For training, this equated to a sample size of 245 data points (71 actual TOD data points and 174 non TOD data points) input to a decision tree algorithm. The visualised A380 decision tree for cluster 1 is shown in Figure 12. To facilitate understanding of the decision tree and its results, the details of each node is as follows (top to bottom): 1) Condition statement 2) gini index, 3) total number of samples that fits into the category, 4) sample break up, [non TOD locations, TOD locations].

A simplified decision tree is also shown in Figure 13. From the A380 - cluster 1 decision tree, the root node is based on the ground speed of aircraft, should ground speed fall below 424 knots, the aircraft was evaluated to be past the TOD point and already in descent flight phase, resulting in a TOD prediction of 0 which equates to no TOD occurrence. If however, ground speed were to be above 424 knots, $TOD_{DIFF}(micro)$ is then considered. TOD occurrence is predicted (TOD = 1) when aircraft was $TOD_{DIFF}(micro)$ was 31.574 NM or less. However, it is not applicable for aircraft above exiting cruise at FL355. A prediction of TOD occurrence in this scenario would only be expected when $TOD_{DIFF}(micro)$ was below 43.079 NM and above 37.017 NM in addition, ground speed was also expected to be above 435.5 knots. This was accepted as departing from higher altitudes would require a longer

Fig. 12. Decision tree diagram for A380 - cluster 1

Fig. 13. Simplified decision tree diagram for A380 - cluster 1

track distance for descent phase resulting in an earlier TOD as compare to lower altitudes. From the given decision tree prediction was predominantly dictated by $TOD_{DIFF}(micro)$ and ground speed variables.

I. Results evaluation and model validation

For flight trajectories relating to cluster 1 of the A380 fleet, a total of 8 flights or 10 percent of flight data was used to evaluate the model. While a higher testing dataset maybe

argued for, training the decision tree model with 90 percent or 71 flights was prioritised in view of small dataset obtained. From the given Decision tree model, a prediction range of +31NM (prior) to -6NM (post) from actual TOD was found, with Ground speed dictating the extent of negative limit. While a prediction estimation of 31NM from actual TOD is comparably large for a prediction model, flight time taken to travel 31NM, was amounted to an estimate of 4 minutes with an average speed of 450 knots (ground speed). As this prediction model is dependent on dataset applied, the 4 minutes prediction window can be related to the variance in TOD within the cluster. Moreover, because $TOD_{DIFF}(micro)$ references the distance to runway threshold instead of the metering fix, a larger margin of error between expected and actual TOD location is anticipated as historical data does not fully conduct CDA, leading to reduced accuracy in TOD prediction. From the given case study, while the use of a substantially larger dataset with higher recording quality is recommended (as per the common practice for ML use in aviation), it is believed that the use of a machine learning method in predicting TOD location can be achieved using prior cruise parameters alone.

V. CONCLUSION

Under the light of increasing air traffic congestion and reduced environmental impacts, the growing need for tailored approaches has arisen. To coordinate such operations, the ability of ground based automation must be improved specifically in the ability to predict TOD of an arriving aircraft. While predictors have been investigated the acquisition and fidelity requirements of data may not always be readily available for ground-based operations. Therefore from this research, an alternative machine learning approach is developed to predict TOD of arrival aircraft using prior cruise parameters alone. In this paper, an ensemble method of DBSCAN-K-means is attempted to determine common TOD clusters and the resultant average expected TOD locations based on the approach direction, fleet type. These macroscopic average expected TOD positions were then compared to the more microscopic pilot's operational perspective using the CDA assumption formula to determine a proximity to expected TOD parameter. From historical data and the employed unsupervised learning model, the derived parameters of proximity to expected TOD, current flight level, approach heading (associated flow direction), ground speed, fleet/aircraft type and landed runway are ingested into a decision tree model for training then for prediction of TOD location of a particular flight. To illustrate the methodology a case study inside the Singapore flight information region (FIR) is conducted with particular interest set on the inbound A380 fleet. From the presented decision tree of the A380 - cluster 1 decision tree, accuracy of the model was given by a prediction range of 31NM (prior) to -6NM (post) range from actual TOD location (equivalent to 4 minutes of flight time). While recommendations for increased data use and inclusion of other significant parameters have been made. This paper has sufficiently shown the possibility of using a machine learning approach to predict TOD of a flight

considering both prior cruise characteristics and operational considerations of pilots alone. An accurate predictive model for TOD location may prove effective in redesigning of STARs and in developing procedures to facilitates CDA in busy terminal airspace. However, to increase the robustness and accuracy of TOD prediction, other factors such as wind and weather data, holdings and altitude/speed restrictions at metering fixes, airline operation policies are recommended for use in future works. While prediction of TOD is crucial in enabling accurate ground based automation for tailored approaches, it is only one part of the puzzle in ensuring efficient yet safe travel.

REFERENCES

- [1] R. Sáez García and X. Prats Menéndez, "Comparison of fuel consumption of continuous descent operations with required times of arrival," in *ICRAT 2020: papers & presentations*, 2020, pp. 1–8.
- [2] P. Pradeep and P. Wei, "Predictability, variability and operational feasibility aspect of cda," in *2017 IEEE Aerospace Conference*. IEEE, 2017, pp. 1–14.
- [3] L. Vempati, "Observed impact of traffic and weather on continuous descent and continuous climb operations," in *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*. IEEE, 2015, pp. 1C5–1.
- [4] E. Itoh and M. Mitici, "Queue-based modeling of the aircraft arrival process at a single airport," *Aerospace*, vol. 6, no. 10, p. 103, 2019.
- [5] S. G. Park and J.-P. Clarke, "Optimal control based vertical trajectory determination for continuous descent arrival procedures," *Journal of Aircraft*, vol. 52, no. 5, pp. 1469–1480, 2015.
- [6] C. M. Johnson, "Analysis of top of descent (tod) uncertainty," in *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*. IEEE, 2011, pp. 2E3–1.
- [7] L. Stell, "Analysis of flight management system predictions of idle-thrust descents," in *29th Digital Avionics Systems Conference*. IEEE, 2010, pp. 1–E.
- [8] A. Maheshwari, N. Davendralingam, and D. A. DeLaurentis, "A comparative study of machine learning techniques for aviation applications," in *2018 Aviation Technology, Integration, and Operations Conference*, 2018, p. 3980.
- [9] K. Mumtaz and K. Duraiswamy, "A novel density based improved k-means clustering algorithm—dbkmeans," *International Journal on computer science and Engineering*, vol. 2, no. 2, pp. 213–218, 2010.
- [10] H. J. D. Reynolds, T. G. Reynolds, and R. J. Hansman, "Human factors implications of continuous descent approach procedures for noise abatement in air traffic control," in *Proc. 6th USA/Europe Air Traffic Management R&D Seminar, Baltimore, MD, June*. Citeseer, 2005.
- [11] A. A. Barsheshat, "Implementation of ads-b systems—benefits and considerations," in *2011 Tyrrhenian International Workshop on Digital Communications-Enhanced Surveillance of Aircraft and Vehicles*. IEEE, 2011, pp. 197–201.
- [12] R. T. C. for Aeronautics, *Minimum Aviation System Performance Standards for Automatic Dependent Surveillance Broadcast (ADS-S)*. RTCA, Incorporated, 2002.
- [13] A. ANDREEVA-MORI and T. UEMURA, "Optimal top of descent analysis in respect to wind prediction errors and guidance strategies," *TRANSACTIONS OF THE JAPAN SOCIETY FOR AERONAUTICAL AND SPACE SCIENCES, AEROSPACE TECHNOLOGY JAPAN*, vol. 15, no. APISAT-2016, pp. a45–a51, 2017.
- [14] "Descent, approach and landing - a guide for controllers," Nov 2012. [Online]. Available: <https://www.skybrary.aero/index.php/Descent,ApproachandLandingAGuideforControllers>
- [15] M. LaMarr, N. Ho, W. Johnson, V. Battiste, and J. Biviano, "Enhancing pilot ability to perform cda with descriptive waypoints," in *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*. IEEE, 2011, pp. 6C3–1.
- [16] U. M. Fayyad and K. B. Irani, "The attribute selection problem in decision tree generation," in *AAAI*, vol. 92, 1992, pp. 104–110.
- [17] Y.-C. Chang, W.-H. Lee, and C.-J. Hsu, "Identifying competitive position for ten asian aviation hubs," *Transport Policy*, vol. 87, pp. 51–66, 2020.