

Evaluation of feasible machine learning techniques for predicting the time to fly and aircraft speed profile on final approach

Predictive dynamic support tool on final approach

F.F.Herrema & V.Treve

Airport Research
EUROCONTROL
Brussels, Belgium

floris.herrema@eurocontrol.int vincent.treve@eurocontrol.int

R.Curran & H.G.Visser

Air Transport and Operations
TU Delft
Delft, Netherlands

r.curran@tudelft.nl h.g.visser@tudelft.nl

Abstract — currently, at many airports, the runway throughput is the limiting factor for the overall capacity. Among the most important constraining parameters is the separation minima expressed in distance. On the top of these minima, the difference of the leader and follower aircraft speed profiles imposes to consider buffer to cope with compression effect. Currently, Air Traffic Control Officers (ATCO's) take these buffers on the basis of their training and experience. However, this experience will not be sufficient to safety deploy advanced concepts, like pair-wise separations, that increase variability in the separations to be delivered and therefore in the compression buffer to be considered. Systematic analysis of years of radar tracks has allowed to better predict the buffers to apply by characterising the time to fly (T2F) given a separation distance and True Airspeed (TAS) profile as a function of meteorological parameters.

This paper presents how Machine Learning (ML) techniques may be used for predicting the T2F and TAS profile on final approach. Different ML techniques will be assessed on their forecast performance, computational time and amount of data needed for delivering a reliable prediction. The techniques will be applied on 2 different major European airports traffic and will be benchmarked against Optimized Runway Delivery (ORD) study using a Model Based Approach (MBA) for deriving the T2F and TAS. As a result the most efficient ML techniques will be applied on two case studies for predicting the T2F and TAS.

Keywords- T2F, TAS, ML, ensemble.

I. INTRODUCTION

ML can be used to identify patterns¹ and to observe 'what-if' scenarios in past data. These patterns can be transferred into 'what-if' statements by analysing relations between the response variables (T2F and TAS) and the prediction variables highlighted in Table 1. This analysis is needed to predict forthcoming operational risks during real time operations like

loss of separation [1]. Such a prediction would feed 'what-if' tools at the airport to alert ATCO about impending aircraft behaviours.

Presently Distance Based Separation (DBS) or Time Based Separation (TBS) rules are applied during final approach. As a next step the 'Dynamic pair-wise separation' concept is proposed to allow controllers to sequence arriving and/or departing aircraft using Time-Based, Weather Dependent and Pair-wise wake turbulence separations. The efficient deployment of such concept needs a reliable prediction of the T2F and TAS, which is mainly influenced by the aircraft type and wind profile. With this respect, it is envisaged to progressively move from a MBA to a ML approach for coping with the variability of aircraft speed behaviours. In this paper, ML techniques will be assessed on their capabilities to produce fast and accurate predictions and their capabilities to test a large number of 'what-if' statements.

This paper presents 10 feasible ML techniques; the amount of data needed, Principal Component Analysis (PCA) and feature analysis for predicting the T2F and TAS profile per wind-band on final approach based on 15 prediction variables. All relevant scenarios between combined data, ML techniques and problems will be assessed. As a result the most efficient ML techniques will be applied on two case studies for predicting the T2F and TAS.

The structure of this paper is as follows; Firstly, the methodology and data sources such as; aircraft data, wind and speed profiles are described. Secondly, the ML context is outlined. Thirdly, the pre-processing steps are elaborated. Fourthly the results of the two case studies and the respective prediction error are outlined in subchapter VI. Finally conclusions and recommendations are drawn.

¹Patterns in this context are defined as the normal distributions (from -2 to +2 sigma) of T2F and TAS for a given flight and for a range covering the last 10NM of the final approach

II. DATA, PREDICTION VARIABLES & MODELLING

In order to predict T2F and TAS profiles on final approach, two complementary sources of information are used; aircraft performance data and weather data.

II.A Aircraft Performance & Weather data

Aircraft Performance is extracted from Radar data and has been provided by the Air Navigation Service Providers (ANSPs) for four airports. For one other airport, ADS-B data has been used. For each airport, the radar data cover 2 months of operations in 2012 or 2013. The ADS-B data covers two weeks of data in 2013. In total, the data comprises about 130,000 flights. For each radar point, the flight ID, aircraft type, actual time, lateral and longitudinal position coordinates, altitude and ground speed is recorded with a 4s refresh rate. The focus of the analysis is on approaching aircraft to single runways over the last 10NM. All variables can be found in Table 1.

The headwind (HW) profile and visibility measurements were gathered from Airport 1 and Airport 2 for respectively a period of 2 months and 3 years. The HW profile is defined by four measurements at heights of 10m, 500m, 1000m and 3000m. The HW profile is analysed per 10 minutes. Wind measurements are grouped into six different wind bands; 0-5kt TW (tailwind) and 0-5kt, 5-10kt, 10-15kt, 15-20kt and 20-25kt (headwind). The HW profile and visibility date and time have been included as input variables into Table 1.

II.B Prediction variables

Table 1 gives all the 15 input prediction variables per 0.5NM segment. However, for some prediction variables the number is constant such as; recatEU, rwy and FAF or some are not considered such as gspass from 10NM till 0.5NM and AC from 19.5 till 10NM. The predictive response variables in our model are the T2F, y , and TAS, y' and are outlined in Table 1 as number 16 and 17.

Table 1. Prediction and response variables

Variable	Description
1.Flightnr	Flight number
2.Apt	Airport
3.actype	Aircraft type
4.dpass	Distance from threshold (NM)
5.hpass	Height from threshold (m)
6.gspass	Ground Speed (kts) from 10 till 19.5NM
7.rwy	Runway (degrees)
8.recatEU	RECAT EU category
9.METARhwind	Headwind (kts)
10.METARcwnd	Crosswind (kts)
11.METARvsby	Visibility (m)
12.ICAOcombi	ICAO combination
13.actypecombi	Aircraft type combination
14.ACin10NM	Number of aircraft between 0 and 10NM
15.FAF	Final approach fix
16.T2F (response)	Time to fly (s)
17.TAS (response)	True Airspeed (kts)

II.C Modelling of ML techniques

The ML techniques will be addressed by modelling the predicted T2F and TAS values under different weather conditions. The T2F is a continuous variable that is computed depending on variables such as aircraft type, airport and parameters related to weather conditions. The approach is based on learning a model per airport and aircraft type - A320 at Airport 1 and the B738 at Airport 2. The results for these two case studies will only be shown for the best 3 feasible ML techniques. Going one step further, we propose to approach the prediction as a multi-task learning problem. This approach can lead to a better model for the main task by exploiting the commonality among the tasks. In this research, this leads to the following consideration: instead of predicting T2F and TAS for each segment of 0.5NM from 0.5NM to 10NM individually, we propose to exploit multi-task learning by predicting the segments altogether. By solving the regression (III.A) problem jointly for all these segments, we expect to improve the performance of the regression compared to the case where the segments are considered independently. The rationale behind this is that although the distribution of the T2F values depend on the segment, the behaviour of the aircraft on all the segments is subjected to the same conditions.

III. CONTEXT-MACHINE LEARNING

This section describes the feasible ML category and techniques for predicting the T2F and TAS profile on final approach.

III.A ML techniques classification [2]

ML techniques can be classified into different categories following three main strands; unsupervised learning, supervised learning and reinforcement learning. Supervised learning can be divided into two different subcategories which are classification and regression. In this study, there will be a primary focus on supervised regression learning since these are often computationally efficient for predicting the T2F and TAS (real numbers) whereas classification is often used for binary predictions such as go-arounds. For supervised regression learning, we propose two approaches that can be considered as baselines for this study [2] [13]. The first method is based on linear regression techniques and the second method on neural networks. These two methods can be divided into 10 sub techniques based on multi-task learning. Multi-task techniques are selected since we try to jointly fit the T2F and TAS for all segments from 0 till 19.5NM (0.5NM step). By definition, a multi-task learning approach learns a problem together with other related problems, all at the same time. Learning multiple related tasks simultaneously has been empirically [3, 4, 5, 6, 7, 8, 9, 10] as well as theoretically [3, 11, 12] shown to often significantly improve performance relative to learning each segment independently. The 10 feasible ML techniques are outlined in subsection III.B and III.C.

III.B Regression techniques to be tested [2]

The regression techniques **fitglm**, **stepwiseglm**, **ridge regression** and **Lasso** are well-suited for this problem. A variant of these techniques, called **Elastic net**, which combines the penalties of both methods and which is also a good candidate to tackle this particular prediction problem [14]. The Mean Square Error (MSE²) serves as cost function for these algorithms. These techniques are applicable in the case of multi-task regression and are referred in the scientific literature as **multitask regularized regression**.

III.C Neural networks techniques to be tested [2]

Approaches based on neural networks are also good candidates for the problem. We recommend the **Multi-Layer Perceptron** (MLP) as a baseline to tackle our case. The loss function used to train the network will be the MSE. Neural networks are widely used in scenarios in multi-task learning, by making use of the fact that the underlying representation of the problem is inherently learnt during the training process. Depending on the amount of data at disposal, deeply connected neural network architectures will also be considered such as; **Auto Encoder**, **Boltzmann** and **Recurrent Neural Networks**.

IV. PRE-PROCESSING

This section describes the pre-processing steps to come up with a usable aircraft performance data set. This data set is needed to train a T2F and TAS prediction model. Each pre-processing step is detailed below:

- A. **Compute T2F and TAS for each sample;** Before feasible ML techniques can be applied first the T2F and TAS profile are extracted for each segment of 0.5NM, 5kts wind band and aircraft type. Remove the samples where the T2F and TAS of one segment is more than 2 standard deviation away from the segment mean. This forms a matrix **Y** where each row represents a flight and each column a segment.
- B. **Feature selection;** The RreliefF technique is applied before a model is learned.
- C. **PCA:** finding out which features are important for best describing the variance in a data set.
- D. **Construct the datasets:** based on different data sources and the Table 1 mentioned variables. Furthermore standardize feature matrix **X**.
- E. **Stability of three different data parts:** split the matrices **X** and **Y** in two subsets **X_{train}**; **Y_{train}**; used to train the model and **X_{test}**; **Y_{test}** used to evaluate the model accuracy. For those experiments the data is split into 70% of training data, 15% of test data and 15% of validation data (standard hold-out).
- F. **Accuracy of data and outliers:** in the last pre-processing step the accuracy is measured and the outliers are shown.

IV.A Compute T2F and TAS profiles

The T2F is computed by the difference in time from a certain distance till threshold. The TAS is calculated by subtracting HW or adding TW of the wind profile from/to the GS profile. This study works with TAS since this gives a better indication of the speed compensations applied per aircraft type. 20.000 flights were extracted from Airport 1 to cover seasonal variations and to have a minimum of 50 measurements per aircraft type, wind-band and 0.5NM segment. Figure 1 shows an example of the TAS, GS and HW profile of an A318 in 10-15kts headwind as a function of distance from the threshold. The T2F and TAS results for 50 different aircraft types can be found in the report [15].

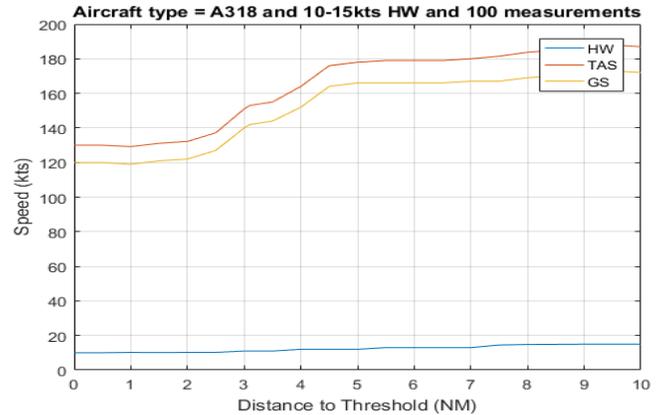


Figure 1. Example TAS, GS and HW profile versus distance to threshold.

IV.B Assessability of feature selection

Before the model will be trained, first the most important (group) features will be selected using PCA and RreliefF modelling (feature selection). The objective of feature selection is three-fold: improving the prediction performance of the predictors, providing faster and more effective predictors, and providing a better understanding of the underlying process that generated the data [16]. RreliefF has commonly been viewed as a feature selection method that is applied in a preprocessing step before the model is learned [17]. The standard RreliefF regression modelling technique has been extensively discussed in many papers [18]. The technique has been applied on 500 low wind (0-5kt) A320 flights for Airport 1 as showed in Figure 2.

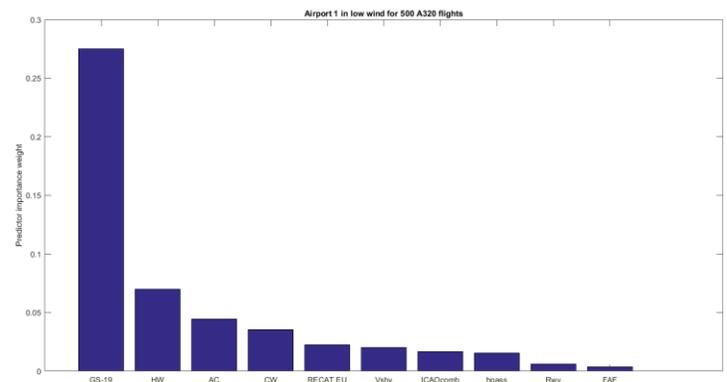


Figure 2. Normalized feature selection using RreliefF algorithm.

²MSE is an estimate of how close to the predicted model mean of your sample mean is likely to be.

Figure 2 shows from left to right the most important succeeding features for Airport 1. The ground speed at 10NM (GS-19) seems to have the most impact on the T2F, followed by the headwind (HW) and Aircraft type (AC). Similar feature relationships are obtained for Airport 2 and different aircraft in low wind. According to the ORD study [19], the top 3 most important theoretical features match with the predicted RreliefF features. Table 2 in section V.A compares the best (group) prediction features for different amount of flights for both PCA and RreliefF.

IV.C Assessibility of PCA

After applying RreliefF (feature selection), PCA will be applied. PCA is a procedure for identifying a smaller number of linearly uncorrelated variables called principal components. The goal of PCA is to show as much of the variability in the data as possible with the fewest number of principal components. The data have been divided into 15 different indicators of aircraft and weather behaviour at 2 different airports, which are showed in Table 1.

Figure 3 shows the top 10 feature selected variables, which are represented in a bi-plot by a vector, and the direction and length of the vector indicate how each variable contributes to the two principal components in the plot. In the new coordinate system, the first axis corresponds to the first principal component, which is the component that explains the greatest amount of the variance in the data, whereby it is obvious that component 2 explains the 2nd greatest amount of variance in the data, etc. In this example, the first principal component, on the horizontal axis, has positive coefficients for GS, Visibility, HW, AC, RECAT CAT, ICAO comb and FAF variables. That is why the seven vectors are directed into the right half of the plot. The largest coefficients in the first principal component are the second, third, fourth and seventh elements, corresponding to the variables HW, GS, RECAT CAT and FAF. The second principal component, on the vertical axis, has positive coefficients for the variables Runway, Height, Cwnd, AC, RECAT CAT, ICAO comb, Visibility, HW and negative coefficients for the GS and FAF variable.

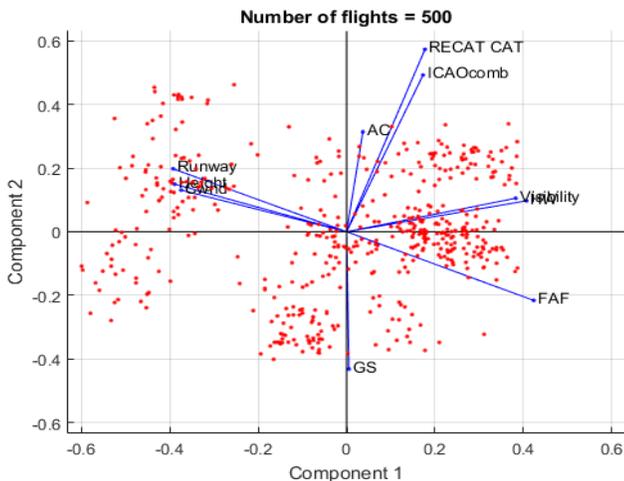


Figure 3. A bi-plot in two dimensions, to find the relation among different variables.

Since figure 3 doesn't explain enough of the variance in the data of the first two principal components, Table 2 and 3 takes also into account component 3 and 4.

IV.D Construct the datasets

The first dataset includes the features flightnr, dpass, hpass, gpass, rwy, RECAT EU, METARvsby, ICAO comb, actypecombi, ACin10NM and FAF. Please note that we only consider the measurements from 19.5NM to 10NM. This forms the feature matrix X where each row is a flight and each column a feature. Another dataset is built from 10NM to 0.5NM with the same features plus the headwind at each segment determined as the difference between ground and indicated air speed. When this is done the historical data will be divided into predictor variables and response variables. Finally, for each column X subtract the columns mean and divide by their standard deviation.

IV.E Stability of three different data parts-cross validation

To check the stability of different data parts, the data will be randomly divided into training, validation and testing subsets. It has been assumed that the default ratios in this study for training, testing and validation are 0.7, 0.15 and 0.15, respectively. The model is adjusted accordingly when training it. The validation is used to measure network generalization, and to halt training when generalization stops improving. To prove that a randomly selected data set is stable, epoch and validation checks are performed. Epoch indicates the amount of a single pass through the entire training set, followed by testing of the verification set. Thereafter we check convergence on the validation and at the end of the learning process the model is evaluated on the test set. The test has no effect on the training and therefore provides an independent measure of network performance during and after training. Figure 4 shows a final approach trained speed model by selecting 5000 A320 flights.

By training the model according to the above described method, a good representation of real life flights will be given and unstable data parts are neglected.

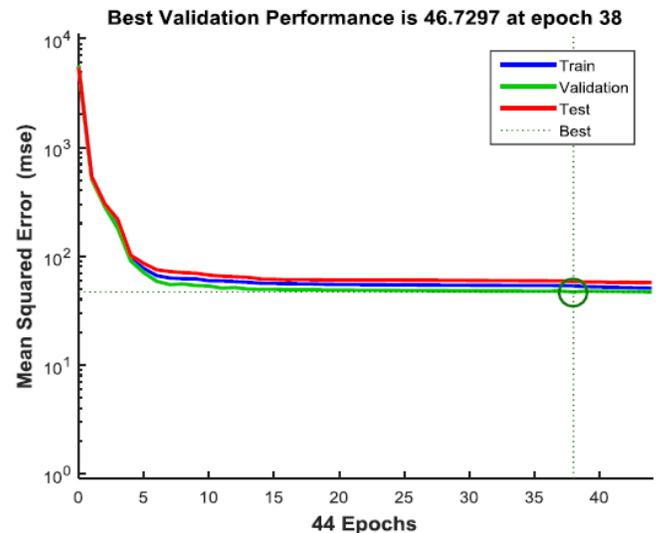


Figure 4. MSE versus amount of epochs for 5000 A320 flights.

IV.F Accuracy of data and outliers

The analyses and detection of abnormal flights and outliers of real flights will be done in the next paper. However, also outliers exist when building predictive models. With outliers we mean when a data point or flight is not consistent with the other data points. One way to show this inconsistency is by plotting the regression for training, validation, test and all. Figure 5 shows such an example where the regression R values measure the correlation between outputs and targets. An R value of 1 means a close relationship, 0 a random relationship.

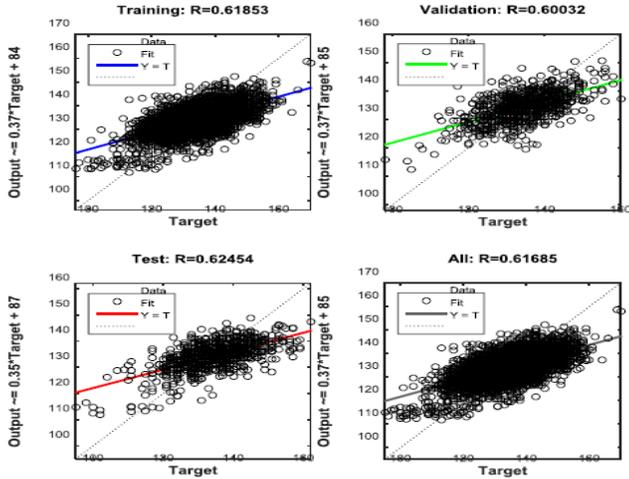


Figure 5. R values for test, training, validation and all for an A320 predictive TAS model

Analysing these graphs shows that there are indeed many outliers. It will be obvious that by neglecting them in the target set, a better R value will be obtained for the predicted model. Doing this for the above example results in an overall R value of 0.69 instead of 0.61 presented in Figure 5.

V. RESULTS

This chapter shows the results of the feasible ML techniques PCA and RreliefF for prediction of the T2F and TAS profile on final approach. The best technique will be assessed on the amount of neurons and minimum amount of flights needed to come up with an accurate prediction model.

V.A PCA groupings and RreliefF for A320 at Airport 1

Using PCA dimension reduction and/or feature selection will automatically not result in a better prediction model. It could happen that by excluding variables, you exclude automatically variables that are correlated with each other. It has been tested if by applying PCA and feature selection before training a ML model result in; (1) less time to compute, (2) a lower Mean Squared Error (MSE) and (3) an increased accuracy (lower sigma). Based on Figure 4 and MBA experiences for different types and wind conditions, analysis are executed on the MSE by excluding expected correlations compared to including them. First Table 2 compares the important (group) prediction features for different amount of flights for both PCA and RreliefF.

Table 2. PCA and RreliefF outcome for different amount of flights

Number of flights	PCA	RreliefF
50	9,6 (group 1) 3,10,11,5,8,12 (group 2) 7,15 (group 3)	6,3,9,10,8, 12,11,15,7,5
100	3,9,6 (group 1) 10,12,8,11,5 (group 2) 15,7 (group 3)	9,6,3,12,10,8,11,15,5,7
300	6 (group 1) 9,3,8,11 (group2) 12,10,5,7 (group 3) 15 (group 4)	6,9,3,10,8,11,12,5,7,15
500	6 (group 1) 9,3,8,11 (group2) 12,10,5,7 (group 3) 15 (group 4)	6,9,3,10,8,11,12,5,7,15

First from this table it can be concluded that, based on PCA, 4 main groupings are correlated with each other. The numbers correspond to a certain prediction parameter and can be found in Table 1. After applying RreliefF we verified with PCA that above 400 flights, the prediction parameters influencing the response for the A320 flights at Airport 1 in low wind remain stable. At this stage the minimum amount of 400 should be inserted for designing an accurate prediction model. Table 3 shows for the MLP the TAS MSE and sigma results of the Table 2 mentioned groupings. These results are obtained by building a predictive model for 500 and 250 A320 flights in low wind at Airport 1 and by in-and excluding group correlations from Table 2. The MSE and sigma results are averaged per 0.5NM segment.

Table 3. MSE and sigma results for 4 different groups

Group	MSE-500	Sigma-500	MSE-250	Sigma-250
Group 1	49.1	6.5	57.7	7.7
Group 1, 2	46.0	6.1	55.1	6.8
Group 1, 2, 3	48.0	6.3	57.2	7.5
Group 1, 2, 3, 4	48.3	6.3	59.5	7.4

Analysing the MSE using all decision parameters compared to the first three groups, first two groups and first group result in respectively; a 1% and 5% improvement and 2% reduction for the TAS MSE (table 3). The same results are obtained for the T2F.

For this A320 flight case we conclude that after PCA we only apply the variables that are correlated with component 1 and 2. This can be explained by the fact that the first two principal components for PCA cover around 90% of the correlation for the response variables. Furthermore the time for learning the model stays the same and the sigma values are lower (increased accuracy) for 500 flights compared to 250.

V.B Assess feasible ML techniques

In this subsection the 10 feasible ML techniques will be assessed on their training time (speed), number of parameters and performance indicators. The most important performance parameter to be minimized by the predictions of the models is the Root MSE (RMSE). The RMSE will be calculated from 8.5NM to 0.5NM according to Equation 1;

$$RMSE_{8.5-0.5NM} = \sqrt{\frac{1}{N} \sum_{i=0}^N (T2F_{8.5-0.5NM} - T2'F_{8.5-0.5NM})^2}$$

Equation 1

Where $T2'F_{8.5-0.5NM}$ is the sum of the predicted T2F for each segment between 8.5NM and 0.5NM before runway. Table 4 shows the outcome for the 10 best feasible regression and neural network techniques. The technique with the highest grade receives 10 points whereby the lowest receives 1 point.

Table 4. Assess feasible ML techniques on different performance indicators

Technique	Performance indicators				Outcome
	Performance (MSE) 3p	Computational time 2p	Apply RreliefF and PCA 2p	Implementation clear decision process 2p	
Lasso	10	9	6	9	78
MLP	5	7	10	8	65
Elastic net	8	5	9	5	62
Ridge	7	8	7	4	59
Auto Enconder	9	4	8	4	59
Recurrent Neural Networks	4	10	2	8	52
Boltzmann	5	3	3	4	35
Regularized regression	2	2	2	2	18
Stepwiseglm	2	2	2	2	18
Fitglm	3	1	1	1	12

From Table 4 we conclude that MLP and Lasso performs best. Both techniques will be combined to design a third feasible technique – ensemble. The third model refers as ensemble which is simply the average of the predictions of the Lasso and MLP. Combining these techniques result in a more robust and accurate ML model [21].

V.C Relation number of hidden neurons versus MSE

Figure 6 shows the MSE outcome versus the number of hidden Neurons for the best performed neural network technique. The MLP outcome has been analysed for 5000, 10000 and 30000 flights in low wind [20]

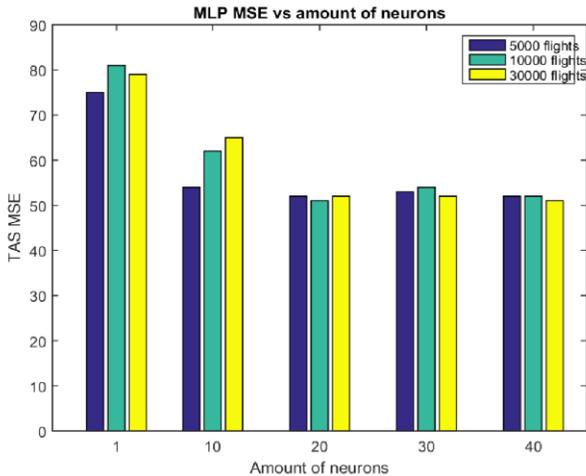


Figure 6. TAS MLP MSE vs number of Neurons

It can be concluded by minimizing the validation MSE, the optimal amount of neurons lies between 20 and 23 for respectively 5000, 10000 and 30000 flights.

V.D Ensemble performance for different number of flights

The ML model is programmed in such a way that it is able to calculate the MSE for different types, wind conditions and for 2 different airports. Figure 7 shows for the ensemble ML technique and MBA the MSE and sigma performances as function of the total number of flights for low (0-5kts) and strong wind (20-25kts) conditions at Airport 1.

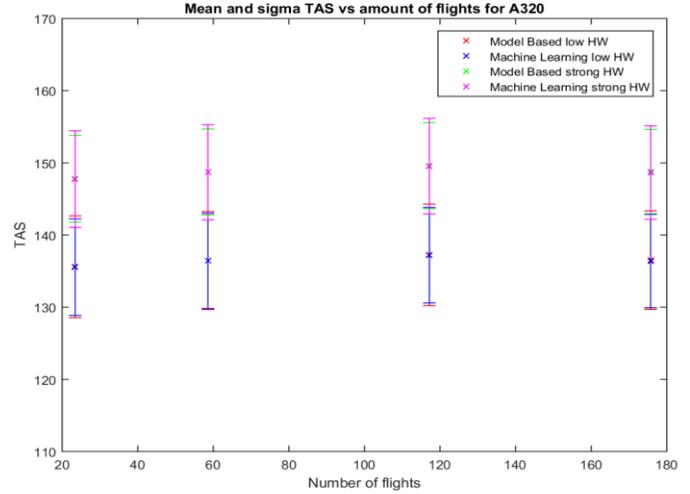


Figure 7. Mean and sigma TAS vs amount of flights for low and strong wind at 1 NM from threshold.

We conclude from Figure 7 that ensemble produces results comparable to MBA (differs between 1 and 2%) and that the standard deviation values are unaffected by sample size. Furthermore by analyzing the MSE, we need 60 flights to build a ML model with accurate results - for the other aircraft types stable MSE values are obtained after learning the model with a minimum of 70 flights. The ensemble model is also validated with an additional data set from Airport 2 and shows comparable results.

Based on the results showed above, it can be concluded that no prediction should be made based on fewer than 60 flights per aircraft type and wind-band. Furthermore outliers like NaN and 0 values should be excluded from the sample data set for valid predictions. The dataset need to be carefully constructed and measured by analyzing the R value for a correct output of the model.

VI. CASE STUDY RESULTS

In this section we analyse two T2F case studies using the Lasso, MLP and Ensemble techniques. During the first case study Airport 1 and aircraft type B738 are analysed. Thereafter we analyse Airport 2 and aircraft type A320. The RMSE of the MBA is estimated using the mean of dataset 1 and dataset 2 (IV.D). We compute the RMSE from 8.5NM to 0.5NM according to equation 1 and accordingly, the RMSE from 4.5NM to 0.5NM which is given by equation 2:

$$RMSE_{4.5-0.5NM} = \sqrt{\frac{1}{N} \sum_{i=0}^N (T2F_{4.5-0.5NM} - T2'F_{4.5-0.5NM})^2}$$

Equation 2

Where $T2'F_{4.5-0.5NM}$ is the sum of the predicted T2F for each segment between 4.5NM and 0.5NM before the runway.

For this airport and aircraft, the estimated errors of the MBA are 6.35 (RMSE 8.5-0.5NM) and 3.65 (RMSE 4.5-0.5NM). Table 5 reports the errors of the different models. Clearly the headwind at each segment helps a lot diminishing the RMSE. Compared to the RMSE of the MBA, none of the model trained on dataset without headwind at each segment achieves the same performance. However, when we also take into account this feature our best model (Lasso) improves by 19.7% over the MBA for the 8.5 to 0.5 NM T2F task and by 19.5% for the 4.5 to 0.5NM task. Averaging the two models lead to a better performance without the headwind but with the headwind, it suffers the bad accuracy of the MLP.

Table 5. Applying top 3 feasible techniques on first case study

Model	Headwind	Size	RMSE 8.5-0.5NM	RMSE 4.5-0.5NM
Lasso	Yes	(1321,331)	5.0	2.9
Lasso	No	(1388,347)	8.0	4.6
MLP	Yes	(1321,331)	5.3	3.1
MLP	No	(1388, 347)	8.0	4.7
Ensemble	Yes	(1321,331)	5.1	3.0
Ensemble	No	(1388, 347)	7.9	4.6

The comparison is not exact but seems to be fair as we compute the RMSE on the same number of segments. Note also that the errors of our models are computed on unseen data. Finally, the bad accuracy of the MLP might be due to the lack of architecture optimization and/or the amount of data.

VI.B Airport 2 and A320

For this airport and aircraft, the estimated errors of the MBA is 4.82 (RMSE 8-0NM) and 3.65 (RMSE 4-0NM). The analysis is the same for this experiment: the headwind at each segment helps diminishing the RMSE. Compared to the RMSE of the MBA, none of the model trained on dataset without headwind at each segment achieves the same performance. When we also take into account the headwind, the MBA is still better than our best candidate by around 2.7% for the 8 to 0 NM T2F task and have the same performance for the 4 to 0 NM task (Table 6). However, the maximum error of our model is lower (on average, all headwind conditions) as it can be seen in Figure 8.

Table 6. applying top 3 feasible techniques on second case study

Model	Headwind	Size	RMSE 8.5-0.5NM	RMSE 4.5-0.5NM
Lasso	Yes	(6753,1689)	4.8	3.8
Lasso	No	(7100,1776)	7.3	4.9
MLP	Yes	(6753,1689)	4.9	3.8
MLP	No	(7100,1776)	7.4	4.9
Ensemble	Yes	(6753,1689)	4.8	3.8
Ensemble	No	(7100,1776)	7.3	4.9

VI.C Absolute error results

The absolute T2F error for the MBA is computed versus the Ensemble method. The comparison is not exact but seems to be fair as we compute the RMSE on the same number of segments (Figure 8). The same has been done for

the TAS. Note also that the errors of our models are computed on unseen data.

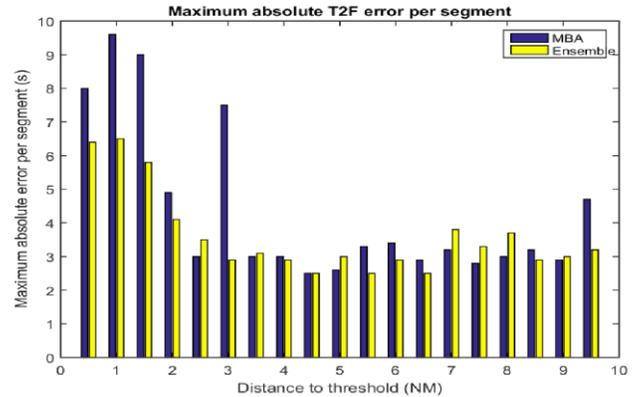


Figure 8. Maximum absolute error per 0.5NM segment

Furthermore the tool is able to calculate for the ML and MBA model for different flight cases (per aircraft type and Airport 1 and Airport 2), the MSE and standard deviation per aircraft type, wind-band and segment.

VII. CONCLUSION

This study assessed feasible ML techniques on their performances for predicting the TAS and T2F. It can be concluded that by using the results of PCA and RreliefF before learning result in a lower MSE, lower sigma and same time compared to the results obtained without using these techniques.

Our experiments show that PCA and RreliefF can discover strong dependencies between attributes, while in domains without such dependencies it performs the same as the MSE. It is also robust and noise tolerant.

Comparing the PCA and RreliefF MSE results using all the decision parameters compared to the first three groups, first two groups and first group result in respectively on average a 1% and 5% improvement and 2% reduction in MSE value for both T2F and TAS (Table 3).

From our experimental results we can conclude that learning multitask regularized regression with RreliefF is promising especially in combination with PCA. RreliefF's good performance and robustness indicate its appropriateness for feature selection.

Ground speed and other information at 10NM together with headwind information seem to capture a lot of the variation of the T2F and TAS in the last 10NM. According to Figure 2, the ground speed at 10NM is the most important feature whereby the headwind vector scores number two.

The multi task techniques Lasso and MLP turned out to be the best feasible and most accurate techniques for predicting the TAS and T2F from 8.5NM till 0.5NM and from 4.5 till 0.5NM. Combining these techniques result in a more robust and accurate ML model which is simply the average of the predictions of the Lasso and MLP - advanced model averaging techniques can be used to enhance the accuracy.

Stable MSE values are obtained when learning minimum 60 flights per aircraft type, wind band and distance from threshold. However when averaging the MSE per 0.5NM segment (10 till 0NM) we suggest a minimum of 400 flights per type and wind band.

Furthermore outliers like NaN and 0 values will be excluded from the sample data set for analysing purposes. The

dataset need to be carefully constructed and measured by analyzing the R value for a correct output of the model.

The ML techniques are more accurate and more robust to changes and they improve in overall over the accuracy of the MBA. We have seen that the standard error decreases with larger sample sizes since the estimate of the population mean improves.

It can be concluded that the optimal amount of neurons for MLP lies between 20 and 23 for respectively 5000, 10000 and 30000 flights in low wind. For high wind values the amount stays the same.

Table 5 and 6 shows that by learning a ML model with HW, the MSE is significantly lower than without HW for both RMSE from 8.5-0.5NM and from 4.5 till 0.5NM. Furthermore the 4.5 till 0.5NM segment has a lower RMSE compared to RMSE from 8.5 till 0.5NM. Finally, the maximum error of our ensemble model is lower compared to MBA.

The results of this study are used as an input by SESAR and EUROCONTROL in the development of a new ATC tool to predict aircraft speed performance. The Leading Optimized Runway Delivery (LORD) tool supports ATCO's to optimize the separation, the buffer and more efficiently and easily deal with the compression effect on the last part of the final approach.

The data supporting the above conclusions was obtained from 2 different airports. To improve verification the results were compared with data from Airport 2 and show significant similarities.

VIII. RECOMMENDATIONS

At this stage the ML tool is able to apply feature selection techniques and ensemble methods for calculating the MSE, standard deviation and amount of measurements for 30 aircraft types, wind-band and 2 different airports. For verification purposes more aircraft performance and weather data per airport should be considered where all airports count the same amount of flights during the same time period. Gathering more data and improving the modelization of the problem might pave the way to a model robust enough to be used in a tactical environment.

Looking at the most important prediction variable – GS at 10NM – might give some operational issues. For predicting the T2F in real life an ATCO has to wait till the aircraft is at 10NM. The T2F for an ATCO is interesting to calculate the compression on final approach using for example the TBS concept. The dynamic TBS for the follower aircraft needs to be known before 10NM. Therefore it is suggested to predict the GS at 10NM of the previous aircraft (based on historical flight information of that time period).

Learn new features such as sequential to visualize the main prediction variables that influence the T2F and TAS. Furthermore find a subspace that captures the variation of the data using PCA dimension reduction.

Learn one task at a time in order to see if the multi-task approach helps and validate that the multi-task approach lead to better results. Learn new ML techniques such as Support Vector Regression (SVR).

In this study the prediction parameters are used from the radar and METAR sets. As a next step the Flight Data Recorder variables will be included for the prediction of the responses, causalities and risks.

A more detailed analysis of the results is needed in order to emphasize the limits of the current approaches. Furthermore, an improved accuracy can be expected from fine tuning of the hyper parameters, network architecture optimization and multiple models averaging.

ACKNOWLEDGMENT

This work is supported by the EUROCONTROL Experimental Centre and the SESAR programme funded by EU, EUROCONTROL and the aviation industry. The authors would like to thank CDG and Vienna airport that provided access to the aircraft flight data. The author also wishes to acknowledge the work of the technical teams that made this work possible.

REFERENCES

- [1] Herrema, F.F. Treve, V. Curran, R.. (2014). Typical additional spacing-buffer to apply at 4DME for delivering separation minima. 2. 8 (2), 100.
- [2] Herrema, F.F. (2015). *Big data analysis and machine learning at airports to support decision making*. Brussels: p10-50.
- [3] R.K. Ando and T. Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *J. Machine Learning Research*. 6: 1817–1853, 2005.
- [4] B. Bakker and T. Heskes. Task clustering and gating for Bayesian multi-task learning. *J. of Machine Learning Research*, 4: 83–99, 2003.
- [5] T. Evgeniou, C.A. Micchelli and M. Pontil. Learning multiple tasks with kernel methods. *J. Machine Learning Research*, 6: 615–637, 2005.
- [6] T. Evgeniou, M. Pontil and O. Toubia. A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. INSEAD N 2006/62/TOM/DS.
- [7] T. Jebara. Multi-task feature and kernel selection for SVMs. Proc. of ICML 2004.
- [8] A. Torralba, K. P. Murphy and W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. Proc. of CVPR '04, pages 762–769, 2004.
- [9] K. Yu, V. Tresp and A. Schwaighofer. Learning Gaussian processes from multiple tasks. Proc. of ICML 2005.
- [10] J. Zhang, Z. Ghahramani and Y. Yang. Learning Multiple Related Tasks using Latent Independent Component Analysis. NIPS 2006.
- [11] J. Baxter. A model for inductive bias learning. *J. of Artificial Intelligence Research*, 12: 149–198, 2000.
- [12] S. Ben-David and R. Schuller. Exploiting task relatedness for multiple task learning. Proceedings of Computational Learning Theory (COLT), 2003.
- [13] Peel, T. (2016). Presentation of the models. In: Peel, T *PoC*. Brussels: d. p8.
- [14] Hui Zou and Trevor Hastie. "Regularization and variable selection via the Elastic Net". In: Journal of the Royal Statistical Society, Series B 67 (2005), pages 301–320 (cited on page 10).
- [15] Herrema, F.F. (2014). *Compression on final approach for optimized runway delivery*. Delft: R. Curran. p58-61.
- [16] Serial metrics. (2013). *Machine Learning*. Available: <http://serialmetrics.com/blog/what-is-machine-learning-feature-selection/>. Last accessed 12-02-2016.
- [17] MARKO ROBNIK S, & KONONENKO, I. (2013). *Theoretical and Empirical Analysis of ReliefF and RReliefF*. Available: <http://link.springer.com/article/10.1023%2FA%3A1025667309714>. Last accessed 12-02-2016.
- [18] MARKO ROBNIK S, & KONONENKO, I.. (2003). An adaptation of Relief for attribute estimation in regression. (3), 9.
- [19] EUROCONTROL (2015). *Optimized Runway Delivery*. Brussels: EUROCONTROL. 10.
- [20] K. Gnana Sheela and S. N. Deepa. (2013). Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Mathematical Problems in Engineering*. 11 (2), p1-11.
- [21] MatLab. (2014). *How can I know training data is enough for machine learning*. Available: <http://stackoverflow.com/questions/24752941/how-can-i-know-training-data-is-enough-for-machine-learning>. Last accessed 28th Feb 2016.