

Predicting Airplane Go-Arounds Using Machine Learning and Open-Source Data [†]

Benoit Figuet ^{1,*}, Raphael Monstein ¹, Manuel Waltert ¹ and Steven Barry ²

¹ Centre for Aviation, School of Engineering, Zurich University of Applied Sciences, 8401 Winterthur, Switzerland; raphael.monstein@zhaw.ch (R.M.); manuel.waltert@zhaw.ch (M.W.)

² Safety and Assurance, Airservices Australia, Canberra, ACT 2601, Australia; steven.barry@airservicesaustralia.com

* Correspondence: benoit.figuet@zhaw.ch

[†] Presented at the 8th OpenSky Symposium 2020, Online, 12–13 November 2020.

Published: 1 December 2020



Abstract: Go-arounds (GAs) are standard air traffic control procedures during which aircraft approach a runway but do not land. The incidence of a GA can subsequently affect the workload of flight crews and air traffic controllers, and might impact an airport runway's throughput capacity. In this study, two different modeling methods for predicting the occurrence of GAs based on open-source Automatic Dependent Surveillance–Broadcast (ADS-B) and meteorological data are presented. A macroscopic model quantifies the probability of a GA within the next hour for an airport by applying a generalized additive model. A microscopic model employs a number of machine learning classifiers on trajectories of aircraft on approach in order to predict if a GA will be performed. Even though the results of the macroscopic model are promising, the information currently available to predict the probability of a GA is not detailed enough to achieve satisfactory predictions. Similarly, the microscopic model is capable of predicting 50% of all GAs, with false positive rate below 7%. Despite the limitations of the quality of the results, the authors are convinced that both modeling methods can be inspiring to other researchers and provide useful insights into the airport system under scrutiny.

Keywords: OpenSky network; go-around; prediction; machine learning; generalized additive model; ADS-B

1. Introduction

Go-arounds (GAs) and missed approaches are air traffic control (ATC) procedures where an aircraft approaches a runway but does not land [1,2]. Common causes for GAs are runways occupied by other aircraft (e.g., due to slow departures) or unstable approaches (UAs), for which the pilots deem the approach not to be within safety margins. UAs are often caused by poor visibility, by unfavorable winds (e.g., crosswinds and tailwinds), or due to approaches flown “too fast, too low, and too close” [3,4].

Internal research at Airservices Australia (the Australian air navigation services provider) has shown that the probability of one GA is 18 times higher within ten minutes of a previous GA. Additionally, the probability of a UA can be accurately measured from energy loss variation [5] and glide slope deviation. Airservices' exploratory work shows the potential of predicting GAs. Furthermore, it highlights the need for further research to determine prediction methods that can be of practical use and that have the potential to be operationalized.

In this paper, methods to predict the probability of GAs are explored. As such, GAs on runway 14 of the Zurich International Airport, Switzerland are predicted. GAs on runway 14 are of particular

interest, since the published GA procedure interacts with the standard instrument departure procedures for runway 16, as illustrated in Figure 1. This interaction, which produced an aircraft proximity event in 2003 [6], is well known to the relevant stakeholders. Subsequently, mitigation procedures have been put in place, which ensure the separation between GAs on runway 14 and departures on runway 16. However, these procedures limit the capacity of the airport. Consequently, predicting GAs on runway 14 could help increase capacity.

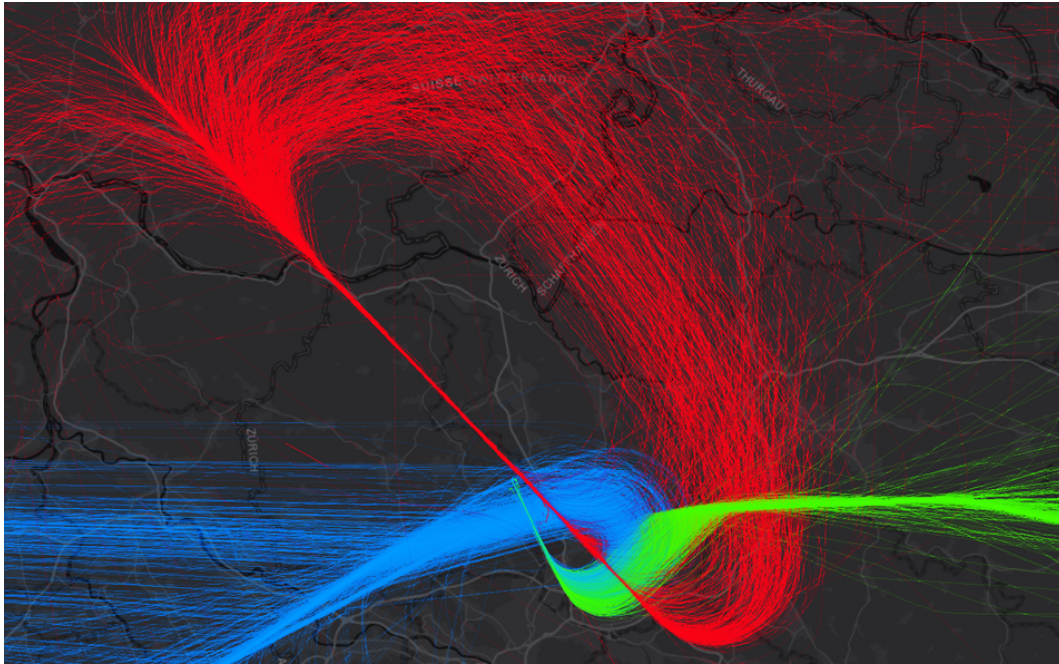


Figure 1. Overview of the situation at Zurich airport. Blue and green: departures on runway 16, red: go-arounds (GAs) on runway 14.

In this paper, two different approaches to predict GAs are presented: one on the macroscopic level and one on the microscopic level. Both methods presented are entirely based on open-source data, namely Automatic Dependent Surveillance–Broadcast (ADS-B) data and meteorological data. The macroscopic model estimates the probability of a GA in the next hour based on information that is available at the time of prediction. To the best knowledge of the authors, such a model has not been presented in the literature. The microscopic model predicts for each landing aircraft at a given cut-off point on the approach whether a GA will be performed after this point or not. A microscopic approach was pursued by Searidge using neural networks to continuously calculate the probability of a GA for each landing aircraft. However, no details about the methods or the performance of the system have been published [7,8].

2. Pre-Processing and Go-Around Detection

In this study, ADS-B trajectories provided by OpenSky Network ([9]) were used. The data cover the period between 1 January, 2017 and 30 June, 2020, and were re-sampled at 5 s intervals. Trajectories may contain outliers and noise, which were subsequently removed with a Savitzky–Golay filter ([10]). The barometric altitude of the observed flights is affected by the atmospheric pressure at the time of observation. This deviation was compensated for by computing the geometric altitude for each landing trajectory using the known elevation of runway 14. Furthermore, the trajectory data were augmented with aircraft data from [11] and meteorological data from [12].

To classify trajectories, a simple and case-specific rule-based approach was used, given its simplicity and robustness:

1. Landings are defined as trajectories, whose last five observations are (i) within a polygon demarcating Zurich Airport’s limits and (ii) below 600 m above mean sea level.
2. Landings on runway 14 are defined as a subset of landings observed at Zurich Airport. To be classified as a landing on runway 14, trajectories (i) stay for at least 5 min within a specifically defined approach corridor (see Figure 2) and (ii) have a heading between 126 and 146 degrees during this time.
3. GAs are defined as trajectories, which (i) first perform an approach, (ii) leave the approach corridor, and subsequently fly for more than 6 min at an altitude above 800 m above mean sea level. For alternative GA classification methods, the reader is referred to Proud [13].

To this end, almost 250,000 landings on runway 14 were identified, of which with 850 were GAs. By visually inspecting the GA data, false positives and outliers (e.g., instrument landing system (ILS) calibration flights, training flights, waiting loops on approach, noisy measurements, etc.) were eliminated, which led to 715 GAs being selected for this analysis (i.e., the GA rate is approximately 3 GAs per 1000 landings).

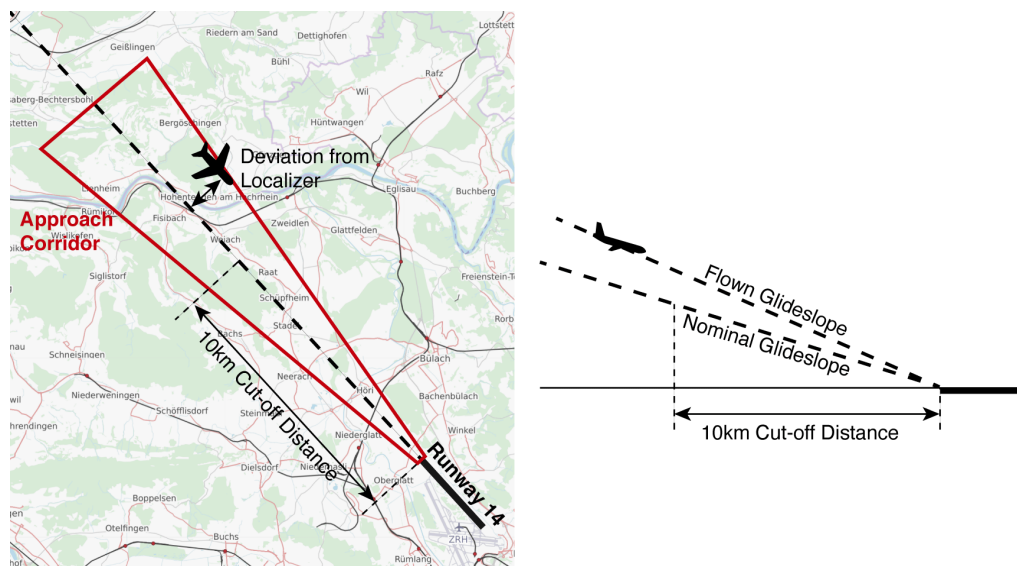


Figure 2. Runway 14 at Zurich Airport and its approach corridor.

3. Macroscopic Model: Prediction of GAs in the Next Hour

The aims of the macroscopic model are twofold. First, it allows valuable insights into the airport system under scrutiny by quantifying the probability of having a GA within the next hour. Second, the model is a helpful tool to facilitate short-term capacity planning for an airport system. For example, a decreased probability of having a GA in the next hour might justify a smaller temporal separation between landing and departing traffic (clearly, a lower probability alone is not enough to decrease the separation between landing and departing traffic; other factors, such as the risk of a wake vortex encounter, must be taken into account as well). The macroscopic model relies exclusively on data that are available at the time of prediction. As such, inputs into the model are the expected traffic of the next hour (e.g., number of landing aircraft, fleet mix, etc.) and meteorological observations derived from METAR data (METAR is a standardized format for reporting meteorological observations at airports) [12].

3.1. Modeling

The probability of having at least one GA in the next hour is modeled with a generalized additive model (GAM). The proposed GAM uses a Bernoulli distribution to model the binary response $y_i \in \{0, 1\}$ or, expressed in practical terms, $y_i \in \{no\ go-around, go-around\}$. The probability of “success”

(i.e., a GA) is typically expressed as being conditional on the predictors as $p_i = P(y_i|X_i) = E[y_i|X_i]$, where the subscript i indicates the i -th observation, p_i is the probability of success of the i -th observation, and $E[\cdot]$ is the expected value. For example, a non-linear logistic regression model can be expressed as

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + f_3(x_{i3}) + f_4(x_{i4}). \tag{1}$$

In Equation (1), the left-hand-side term is called the logit or log-odds function of p_i , and links the predictor $\eta_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + f_3(x_{i3}) + f_4(x_{i4})$ to the conditional expectation p_i . The predictor η_i is an additive model that can contain both linear and non-linear terms (with the subscript j). In Equation (1), the linear terms use parameters β_j , while the non-linear terms are expressed with smooth functions $f_j(x_j)$. The flexibility to model a predictor’s effect on the dependent variable as an additive model makes GAMs ideal for data exploration and for finding suitable model structures. Similarly to linear regression models, GAMs provide statistics on the significance of predictors and other metrics. GAMs are fitted by penalized maximum likelihood. For further information on GAMs, the interested reader is referred to [14,15].

It is worth mentioning that maximum likelihood logistic regression yields asymptotically unbiased estimates. In case of imbalanced data, this assumption might not hold because of the “finite sample and rare events bias” [16]. Particularly, this is problematic when only few observations of the rarer class are available. In this study, 819 h were observed during which at least one GA was performed, out of 15,791 h during which landings on runway 14 were conducted. Even though only about 5% of the observed operational hours contain GAs, the absolute number is assumed to be sufficient according to the simulation studies in [16].

The macroscopic model is based on environmental and aircraft-related predictors that are assumed to be knowable in advance (see Table 1). The GA prediction was made every full hour for the periods where at least one landing was observed on runway 14. The environmental predictors used in the model are based on the last METAR observation available before the prediction, usually at 10 min before the hour. Aircraft-related predictors consist of the number of landings in the next hour as well as an indicator describing the observed probability of a GA for a group of aircraft types (see Table 1).

A suitable model for predicting the probability of having at least one GA in the next hour is:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 \cdot \text{thunderstorm}_i + \beta_2 \cdot \text{acTypes}_{\text{high}_i} + \beta_3 \cdot \text{acTypes}_{\text{uk}_i} + \beta_4 \cdot n_{\text{homecarrier}_i} + f_1(n_{\text{landing}_i}) + f_2(\text{vis}_i) + f_3(\text{crosswind}_i) + f_4(\text{headwind}_i). \tag{2}$$

Table 1. Predictors and their names in the model formula.

Formula Name	Predictor
thunderstorm	Binary factor, true if a thunderstorm was observed ($\text{thunderstorm} \in \{0, 1\}$)
acTypes _{high}	Number of landings of aircraft types classified as high GA probability in the next hour as a fraction of the total number of landing aircraft ($\text{acTypes}_{\text{high}} \in [0, 1]$)
acTypes _{uk}	Number of landings of aircraft with unknown GA probability landing in the next hour as a fraction of the total number of landing aircraft ($\text{acTypes}_{\text{uk}} \in [0, 1]$)
$n_{\text{homecarrier}}$	Number of landings of home carriers in the next hour ($n_{\text{homecarrier}} \in \mathbb{N}$)
n_{landing}	Number of landings in the next hour ($n_{\text{landing}} \in \mathbb{N}$)
vis	Visibility in meters
crosswind	Absolute crosswind component in knots
headwind	Headwind component in knots; positive indicates headwind, negative indicates tailwind

3.2. Results

The estimated coefficients for the linear terms, which can be interpreted in the same way as in a linear regression model, are shown in Table 2. For example, the model predicts, all else being equal, that the presence of a thunderstorm in the vicinity of the airport increases the odds of having a GA by a factor of $\exp(0.95) = 2.6$. The non-linear terms can also provide useful insights into the system and are shown in Figure 3.

Table 2. Linear predictors and their estimated values.

Formula Name	Estimate
β_0 (Intercept)	−3.2
β_1 (thunderstorm)	0.95
β_2 (tc_{high})	0.68
β_3 (tc_{uk})	0.23
β_4 ($n_{homecarrier}$)	0.013

For example, the model predicts that the effect of headwind on the probability of having a GA is highly non-linear. It can be seen that the probability of a GA increases substantially for tailwinds (negative headwind) exceeding 4 knots. Similarly, a crosswind of up to about 5 knots does not affect the probability of a GA, but that probability increases non-linearly after that. The interpretation of the effect of the number of landings is less clear, since the confidence intervals are quite large. These large intervals, particularly towards the end of the scale, indicate that the fitted smooth might not be correct. In this case, this is due to the fact that only a few observations exist with such high numbers of landings on runway 14.

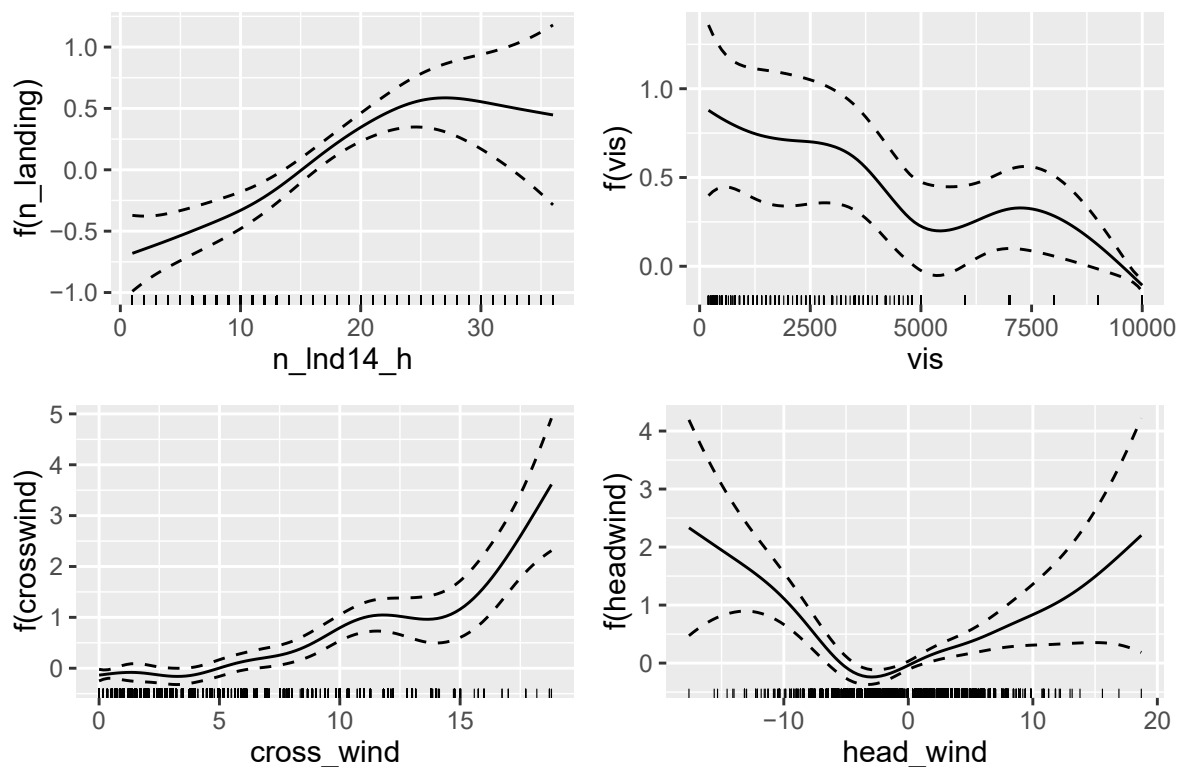


Figure 3. Smooth terms; the solid line is the fitted smooth and the dashed lines are the 95% confidence intervals.

Figure 4 shows the predictions of the model for both GAs and landings. Unfortunately, the predicted probability distributions do not differentiate the landings from the GAs well enough for practical applications. This result was expected to some extent, as the input data used are highly

aggregated and do not contain enough information. For example, the model does not know if the weather reported in the METAR will change during the next hour. It would be possible to increase the fidelity of the model by introducing additional predictors; for example, as suggested in [4], delays and the amount of ground traffic, or the Terminal Aerodrome Forecast (TAF) section of the METAR, which indicates the future trends of meteorological conditions. However, the authors do not expect a significant improvement to the model. Instead, in the next section, a different approach is proposed.

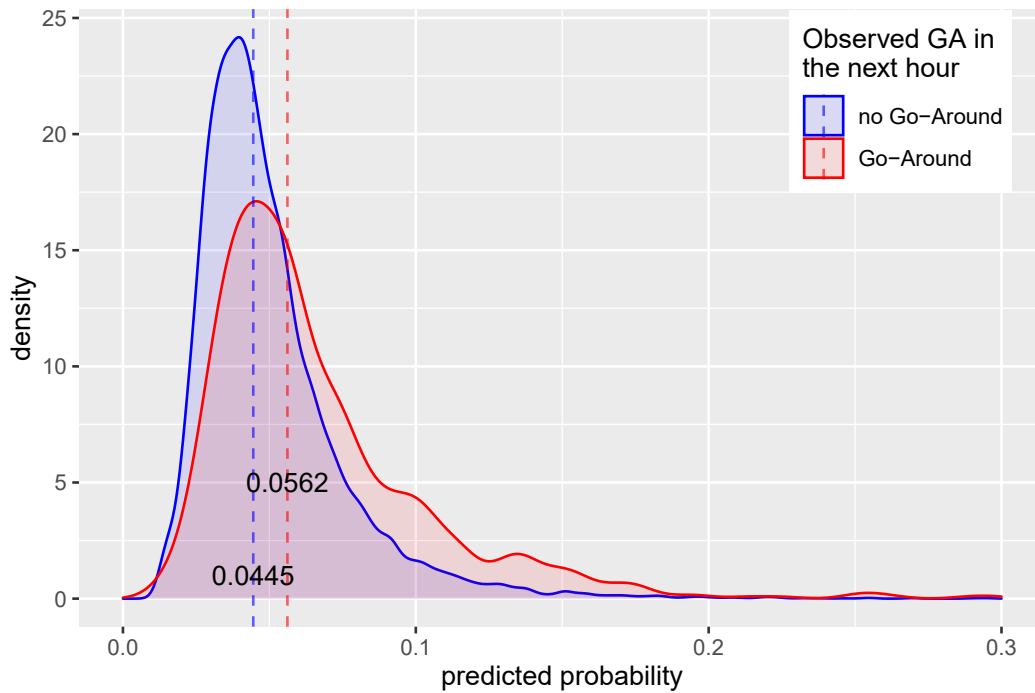


Figure 4. Predicted probabilities. Red: distribution of the predictions when no GAs occurred. Blue: distribution when at least one GA occurred. Dashed: median.

4. Microscopic Model: Prediction of the GA Probability for an Approaching Aircraft

The aim of the microscopic model is to estimate, for each approaching aircraft, the probability of a GA before it is initiated. As such, this information is highly valuable to air traffic control, as potential conflicts can be better anticipated and operation is more efficiently managed.

Following a method presented by Dai, the probability of a GA is estimated at a ‘cut-off point’ [17], which, in this paper, is selected to be a point on the extended centerline of runway 14, located 10 km in front of the threshold. By applying the defined cut-off point, 97% of all observed GAs (695 trajectories) are selected for use in the microscopic model (see Figure 5). Subsequently, for each observed trajectory labeled as an approach to runway 14, the 10 last observations (at a sample rate of 5 s, 10 observations correspond to 50 s of trajectory data) in the approach corridor prior to the cut-off point are used in the microscopic model.

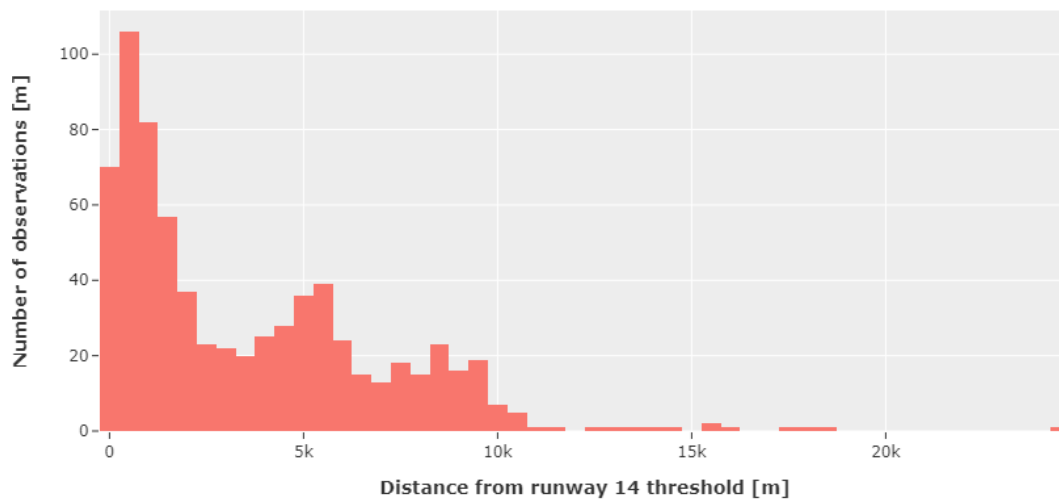


Figure 5. GA initiation distances from runway.

4.1. Feature Engineering

The microscopic model requires additional features, which have to be engineered from the data. These features are (i) a number of stability metrics relating to the stability of an aircraft on the approach, (ii) information on the lead–trail relationship of two aircraft that are approaching simultaneously, and (iii) environmental and aircraft-related information.

4.1.1. Stability Metrics

To describe the stability of an approach, three different metrics (see Table 3) are used: (i) the flown glideslope angle (the glideslope refers to the vertical guidance of the instrument landing system (ILS); for instance, runway 14 at Zurich Airport is equipped with an ILS with a 3° glideslope angle), (ii) the deviation from the localizer (the localizer refers to the lateral guidance of the ILS), and (iii) the specific potential and kinetic energy of the aircraft. The flown glideslope and localizer deviation are depicted in Figure 6.

Table 3. Summary of stability metrics.

Formula	Feature
$\alpha_{gs} = \arctan \frac{h_{ar}}{d_{thr}}$	Glideslope angle α_{gs} , with aircraft at a height above runway h_{ar} and at a distance d_{thr} from the threshold
$\delta_{cl} = dist(airplane, centreline)$	Localizer deviation δ_{cl} : 2D distance between airplane and centerline
$E_s = h + \frac{V^2}{2g}$	Aircraft specific energy E_s (also known as energy height), with aircraft at a height h , ground speed V , and the gravity constant g [3]

Figure 6 shows the profiles and the mean value distributions of the used stability metrics α_{gs} , δ_{cl} , and E_s for a subset of 1000 randomly chosen landings and the 695 identified GAs. Despite an overlap of the distributions of GAs and nominal landings, some GA observations have the tendency to have higher values. Consequently, this indicates that the proposed stability metrics might be useful predictors for unstable approaches.

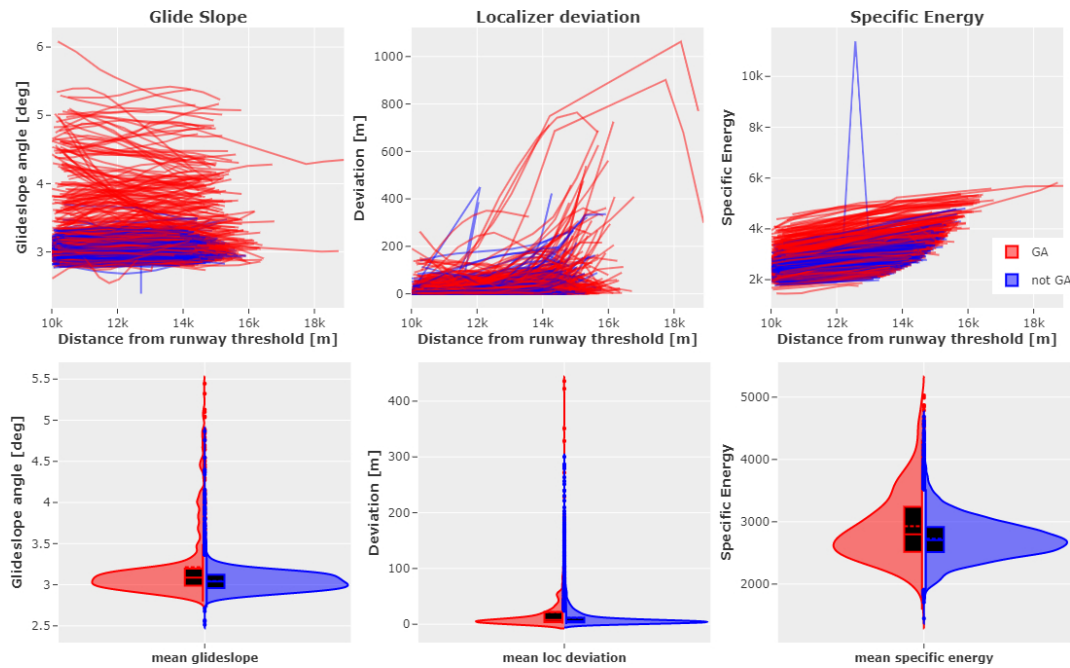


Figure 6. Profile of stability metrics and mean distributions.

4.1.2. Lead–Trail Relationship

When two aircraft are simultaneously approaching the same runway, ATC must ensure spatial separation between the airplanes at all times. Dai [17] suggests that the probability of a GA is substantially increased if (i) there is a loss of separation, and (ii) the closing rate between the leading and the trailing aircraft is “too high”. To model the lead–trail relationship, the time to leader t_{TL} is used in this paper, which is defined as: $t_{TL} = \frac{\Delta_{TL}}{\Delta_{V_{TL}}}$, where Δ_{TL} is the relative distance and $\Delta_{V_{TL}}$ is the relative speed between the leader and the trailer. Table 4 shows that minimum time to leader observed during the approach has a positive impact on GA rates. Additionally, the authors expect the wake category of the leading and trailing aircraft to have an impact on GA rates. Subsequently, both minimum time to leader as well as wake categories are used as features in the microscopic model.

Table 4. Time to leader and observed GAs.

t [s]	Number of GAs observed when $\min(t_{TL}) < t$	$P(GA \min(t_{TL}) < t)$ per 1000 landings
200	260	3.12
180	229	3.47
160	178	3.96
140	126	5.20
120	78	8.99
100	43	23.4
80	16	82.9

4.1.3. Environmental and Aircraft-Related Information

The results of the macroscopic model suggest that weather-related features such as visibility, wind, or the presence of thunderstorms in the vicinity of the airport affect the probability of a GA. Additionally, aircraft-related information, such as the aircraft type, the GA category (the GA category was derived from the data by estimating the probability of a GA for each aircraft type), and a binary factor identifying whether the airline operating the flight is a home carrier (i.e., has Zurich as its home base), are used in the microscopic model.

4.2. Modeling

For the microscopic model, a number of different classical machine learning (ML) classifiers were trained and the corresponding results were compared with each other. The heavily imbalanced data (approximately 3 GAs per 1000 landings on runway 14) constitute a major challenge for the application of classical ML classifiers. As such, Chawla [18] presents several sampling strategies and ensemble-based methods to deal with imbalance. In this project, a downsampling of the majority class (i.e., the landings) has been applied. The data used to train the ML classifiers consist of 1000 random samples of the *non* – *GA* set and all 695 observed *GA* samples. This approach introduces a bias into the predicted probabilities. However, since the microscopic model is used solely for classification of trajectories into the sets $y_i \in \{no\ go-around, go-around\}$, a correction of the bias (e.g., as presented in [19]) is not applied. In order to decorrelate the extracted model features (see Table 5), a principal component analysis (PCA) transformation was applied (see [20]) to the data used for the ML classifiers. In this paper, the following ML classifiers provided in the Scikit-learn package [21] were used: Linear Support Vector Machine, Round Based Function Support Vector Machine, Random Forest, AdaBoost, Naive Bayes, Quadratic Discriminant Analysis, and Gradient Boosting Classifier. Indeed, the classifiers were trained with 50% of the downsampled data, while the other 50% was used for testing and validation purposes.

Table 5. Features used for training of the microscopic model.

Feature Name	Description
mean_glideslope	Standardized mean glideslope
mean_loc_deviation	Standardized mean deviation from localizer
mean_specific_energy	Standardized mean specific energy
min_time_to_leader	Minimum observed time to leader, dummy value if no leader
leader_wtc_cat	Wake turbulence category of leading aircraft
trailer_wtc_cat	Wake turbulence category of trailing aircraft
thunderstorm	Binary factor, true if a thunder storm was observed
ga_cat	Typecode category for GA probability
is_homecarrier	Binary factor, true if the flight is operated by a home carrier
vis	Standardized visibility in meter
crosswind	Standardized crosswind component in knots
headwind	Standardized headwind component in knots

4.3. Results

Figure 7 shows the receiver operating characteristic for each trained model applied on all of the data for runway 14, covering 242,959 approaches. The true positive rate is indicated on the vertical axis, while the false positive rate is displayed on the horizontal axis for different classification thresholds. It can be seen that the Random Forest classifier is the best performing, since it manages to detect half of the GAs with a false positive rate below 7%. While this result seems to be rather good, the number of nominal landings being much higher than the number of GAs, only 2% of the predicted GAs are actual GAs observed in the data.

Figure 8 depicts a comparison of the GA initiation distance with respect to the runway threshold for GAs the model manages to predict (true positives, displayed as yellow-green) and (b) for the ones that have not been predicted by the classifier (false negatives, displayed as red). It can be inferred that the model does not perform well at predicting GAs that are initiated close to the runway threshold, while it performs better at predicting GAs that are initiated further away than 3 km from the threshold. False negatives can be explained by information not present in the model, such as ATC intervention or runway occupancy. Additionally, Figure 9 shows the stability metric profiles and mean distributions for false positives and true negatives, which can be explained by the following factors. First, there is a high chance that the flight crew is capable of stabilizing an unstable approach identified at a cut-off

point located 10 km from the runway threshold. Second, noisy or inaccurate data might contribute to abnormally high approach metric values.

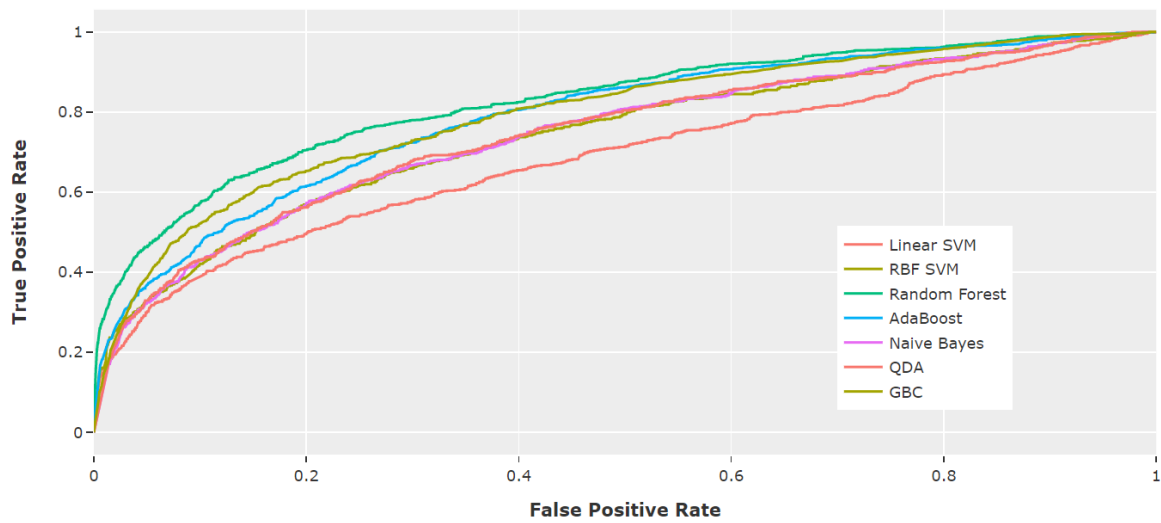


Figure 7. Receiver operating characteristic for GA prediction on all of the data.

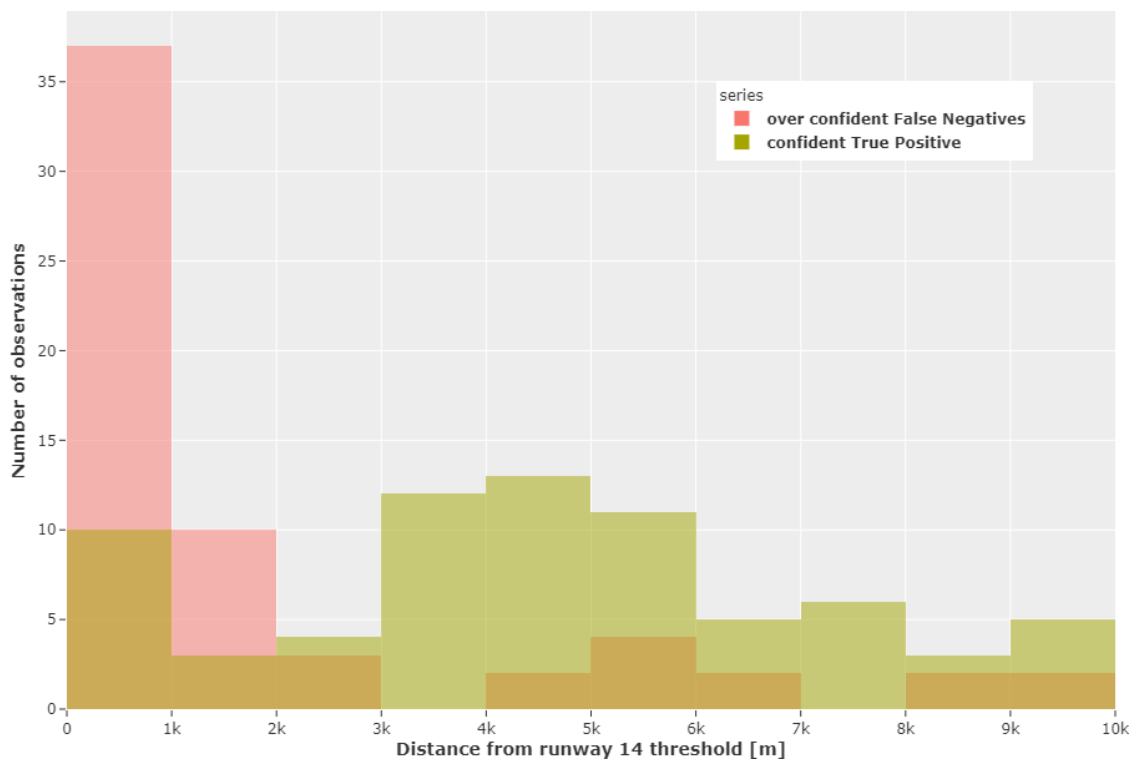


Figure 8. Distance from runway at which GAs are initiated for confident false negatives and confident true positives.

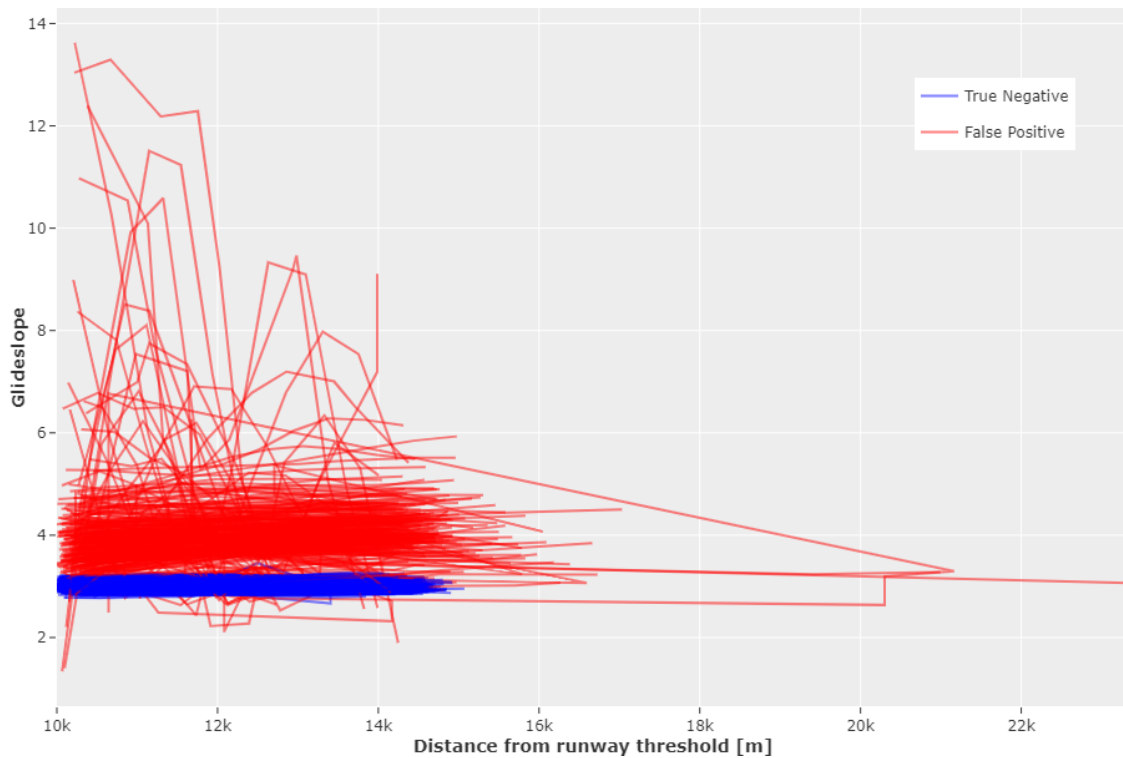


Figure 9. Glideslope for the subset of “true negative” (blue) and “false positive” (red) predictions.

5. Conclusions

In this document, a microscopic and a macroscopic approach for determining the probability of a GA are presented. The macroscopic model aimed at estimating the probability of a GA within the next hour based on information that is available at the time of prediction, while the microscopic model predicted the probability of a GA for each landing aircraft at a cut-off point on the extended runway centerline, located 10 km in front of the threshold. Even though neither of the models performed well enough for real-world applications, they both provide valuable insight into the system. The macroscopic model can be used to understand the nonlinear relation between environmental and operational features and GA occurrences, while the microscopic model could be used as a basis for the development of a tool to assist ATC in identifying potential GAs. A possible extension of this work is the modeling of the probability of a GA not for a single cut-off point, but rather for a large number of points covering an entire section of the approach corridor. Even though the results presented in this paper are somewhat non-conclusive, the proposed methods might inspire other researchers in the area of GA prediction.

Author Contributions: Conceptualization, B.F. and R.M.; methodology, B.F. and R.M.; validation, B.F., R.M., M.W., and S.B.; data curation, B.F. and R.M.; writing—original draft preparation, B.F., R.M., and S.B.; writing—review and editing, M.W. and S.B.; visualization, B.F., R.M., and M.W.; supervision, project administration, acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Swiss Federal Office of Civil Aviation grant number SFLV 2016-072.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CANSO. Unstable Approaches: Air Traffic Control Considerations. Available online: <https://ssd.dhmi.gov.tr/Documents/Unstable%20Approaches-%20Air%20Traffic%20Control%20Considerations.pdf> (accessed on 30 October 2020).
2. Coker, M. Why and When to Perform a Go-Around Maneuver. *Boeing Edge* **2014**, *2014*, 5–11.

3. Gluck, J.; Tyagi, A.; Grushin, A.; Miller, D.; Voronin, S.; Nanda, J.; Oza, N. Too Fast, Too Low, and Too Close: Improved Real Time Safety Assurance of the National Airspace Using Long Short Term Memory. *AIAA Scitech 2019 Forum* **2019**, doi:10.2514/6.2019-0400.
4. Gariel, M.; Spieser, K.; Frazzoli, E. On the Statistics and Predictability of Go-Arounds. *arXiv* **2011**, arXiv:1102.1502.
5. De Boer, R.J.; Coumou, T.; Hunink, A.; Van Bennekom, T. The Automatic Identification of Unstable Approaches from Flight Data. In Proceedings of the Sixth International Conference on Research in Air Transportation (ICRAT 2014), Istanbul, Turkey, 26–30 May 2014.
6. BFU. *Schlussbericht des Büros für Flugunfalluntersuchungen; Schlussbericht 1868; Eidg. Departement für Umwelt, Verkehr, Energie und Kommunikation: Bern, Switzerland, 2005.*
7. Marco Rueckert, Searidge Technologies. Artificial Intelligence in ATM. 2017. Available online: <https://emerj.com/ai-sector-overviews/artificial-intelligence-for-atms-6-current-applications/> (accessed on 11 October 2020).
8. Christian Thurow, Searidge Technologies. Developing and Deploying AI in Risk Adverse Industries. 2017. Available online: <https://on-demand.gputechconf.com/gtcdc/2018/pdf/dc8161-developing-and-deploying-ai-in-risk-averse-industries.pdf> (accessed on 11 October 2020).
9. Schäfer, M.; Strohmeier, M.; Lenders, V.; Martinovic, I.; Wilhelm, M. Bringing up OpenSky: A large-scale ADS-B sensor network for research. In Proceedings of the 13th International Symposium on Information Processing in Sensor Networks IPSN-14, Berlin, Germany, 15–17 April 2014; pp. 83–94.
10. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
11. OpenSky Network. Aircraft Database. 2020. Available online: <https://opensky-network.org/aircraft-database> (accessed on 17 August 2020).
12. Iowa State University. ASOS-AWOS-METAR Data Download. 2020. Available online: <https://mesonet.agron.iastate.edu/request/download.phtml> (accessed on 28 July 2020).
13. Proud, S.R. Go-Around Detection Using Crowd-Sourced ADS-B Position Data. *Aerospace* **2020**, *7*, 16.
14. Hastie, T.; Tibshirani, R. *Generalized Additive Models*; Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis: Oxfordshire, UK, 1990.
15. Wood, S.N. *Generalized Additive Models: An Introduction with R*, 2nd ed.; Chapman and Hall/CRC texts in statistical Science; Chapman and Hall/CRC: London, UK, 2017.
16. King, G.; Zeng, L. Logistic Regression in Rare Events Data. *Polit. Anal.* **2001**, *9*, 137–163. doi:10.1093/oxfordjournals.pan.a004868.
17. Dai, L.; Liu, Y.; Hansen, M. Modeling Go-around Occurrence. In Proceedings of the Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019), Vienna, Austria, 17–21 June 2019.
18. Chawla, N.V. Data Mining for Imbalanced Datasets: An Overview. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2010; pp. 875–886.
19. Saerens, M.; Latinne, P.; Decaestecker, C. Adjusting the Outputs of a Classifier to New a Priori Probabilities: A Simple Procedure. *Neural Comput.* **2002**, *14*, 21–41.
20. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.
21. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).