The Automatic Identification of Unstable Approaches from Flight Data

Teun Coumou, Alexander Hunink & Robert J. de Boer
Aviation Academy
Amsterdam University of Applied Sciences
The Netherlands
rj.de.boer@hva.nl

Abstract — **Unstable approaches have been identified as a major risk factor in approach and landing accidents and runway excursions, but hardly ever lead to go-arounds despite strong safety initiatives. This study challenges the current industry standard for the identification of unstable approaches, as defined by the Flight Safety Foundation Task Force for Approach and Landing Accidents. Based on two independent sets of flight data for 30 approaches, a new algorithm to identify genuinely unstable approaches is designed and validated. This algorithm has been applied at the target airline to better understand pilot decision making in an unstable approach. The adoption of this algorithm to better target the risks associated with unstable approaches is advocated.**

Keywords: Unstable approach, go-around, pilot decirion

I. INTRODUCTION

Over the last ten years, the approach, landing and go-around flight phases account for the largest part of aircraft accidents. In 2011, 63 accidents (68% of all accidents) in commercial aviation occurred during these phases of flight [1], [2]. Unstable approaches are relatively infrequent, amounting to less than 5% of all approaches worldwide, but in nearly all cases the approach is continued making this the leading risk factor in approach and landing accidents and the primary cause of runway excursions during landing [1], [3]. Therefore, the decision to execute a go-around if an approach is not sufficiently stable is encouraged in the interest of safety [4]–[6], but in practice less than 5% of the unstable approaches actually leads to a go-around [5], [6].

To reduce the number of unstable approaches and to encourage go-arounds under these conditions, airlines typically evaluate the flight data retrieved from the aircraft after every flight [7], [8]. The flight data is analyzed for breaches of the stable flight criteria (as detailed in the next section), and for selected flights that are "genuinely unstable" the pilots are invited to discuss the flight progression and the decision not to execute a go-around with safety staff. The selection of these flights requires quite some effort by a flight data analyst and a check pilot, as the currently available algorithms are not able to differentiate between unstable approaches according to the conventional definition and genuinely unstable approaches.

The research presented here was conducted with flight data from an airline that mainly services holiday destinations. Due to the local circumstances at these destinations, a high number of unstable approaches were flagged according to the conventional definition. This resulted in an overload for the flight data analysts to identify those approaches where further analysis and a discussion with the flight crew was warranted. Through discussions with partner airlines it arose that this situation was not a-typical.

A. Problem Statement

The aim of this research is to create a reliable algorithm for use by the airline to identify approaches from flight data that are considered sufficiently unstable by safety staff to warrant further analysis and a discussion with the flight crew ("genuinely unstable approaches").

B. Literature

The Flight Safety Foundation (FSF) Task Force for Approach and Landing Accidents (ALAR) was created in 1996 to support the reduction in aviation approach-and-landing accidents, including those resulting in controlled flight into terrain. The task force has developed recommendations and tools that are made available to the industry [4]. One of its products is the definition of a stable approach, based on the achievement of stability at 1000 feet above airport elevation (instrument meteorological conditions) or 500 feet (visual meteorological conditions). At this point (so-called stabilization gates [9]), the aircraft (1) shall be on the correct flight path; (2) requires only small changes in heading/pitch to maintain the correct flight path; (3) has not less than the correct speed (V_{REF}) and not more than 20 knots more; (4) is in the correct landing configuration; (5) has a vertical speed of no greater than 1,000 feet per minute unless a different rate is required for the approach and a special briefing has been conducted; (6) has an appropriate power setting for the aircraft configuration and is not below the minimum power setting for approach as defined by the aircraft operating manual; (7) all briefings and checklists have been conducted; (8) Specific types of approaches are stable if they also fulfill the following: instrument landing system (ILS) approaches must be flown within one dot of the glide slope and localizer; a Category II or Category III ILS approach must be flown within the expanded localizer band; during a circling approach, wings should be level on final when the aircraft reaches 300 feet above airport elevation; and (9) unique approach procedures or abnormal conditions requiring a

deviation from the above elements of a stable approach require a special briefing. An approach that becomes unstable below the stabilization gate requires an immediate go-around [5]. These criteria have become the industry standard and airlines have generally incorporated them in their standard operating procedures [10]. Based on these criteria, the number of unstable approaches worldwide is estimated to be around 4% of all approaches, down from 8% in 2010. [8], [11].

The continuation of an unstable approach in forfeit of a go-around is proposed to have dire consequences. A study in 1999 found that unstable approaches and the failure to initiate a go-around contributed to 73.5 % of all approach and landing accidents and serious incidents in the years 2003 to 2011 [12], [13]. Boeing identified that for 9 of the 29 landing overrun events which occurred from 2003 to the present the approaches were unstable [14]. However, in practice an unstable approach hardly ever results in a go-around. It is estimated that in about 3% of the unstable approaches a go-around is initiated [3]. In half of these cases, the go-around is initiated at a lower altitude than the prescribed altitude, contributing to a hazardous go-around outcome in nearly 10% of the cases [6], [9]. Efforts by industry safety leaders have been made to stimulate go-arounds in the case of an unstable approach (e.g. [6], [15], [4], [5], [14]).

However, pilots seem to make their own decision based (at least in part) on their perception of risk. The Presage Group [3] found that on an unstable approach the expected braking action had a particularly large impact on whether a go-around was initiated, and further that "pilots' thresholds for calling go-arounds varied as a function of both height above ground level and the instability parameter they were considering as a reason to go around". Wischmeyer contends the commonly accepted high correlation between unstable approaches and bad landing outcomes [16]. He states that: "multiple independent sources demonstrate that almost no unstable approaches end catastrophically, and thus it is inappropriate to consider 'unstable approach' as a causal factor. Rather, 'unstable approach' is almost always correctable, and/or a symptom of other phenomena". He suggests that although a stable approach may seem 60 times safer than an unstable approach for runway overruns, this same data will give a false alarm 49,999 times out of 50,000.

The standard operating procedures of the airline under study have incorporated the criteria for a stable approach as presented above, specifying that a go-around is to be initiated if the criteria are not met at the stabilization gate or cannot be maintained thereafter. The fact that there is room for a personal assessment of the approach is also demonstrated by the fact that the flight data analysts use the standard operating procedures as a guide rather than as normative conditions when deciding whether approaches are unstable or not. The flight data analysts do not have the capacity to discuss all flights that violate the criteria for a stable approach (> 10% of the total), and so priority is given to those approaches that are "genuinely unstable". The personal assessment is further necessitated by the fact that no reference is made to the extent or duration of the violation, or the combination of criteria that are breached.

C. Possible algorithms

To detect genuinely unstable approaches multiple algorithms can be devised. It seems logical to base oneself on the work of the FSF ALARP Task Force and the airline's standard operating procedures. This has led to two possible algorithms. Furthermore, literature on unstable approaches has identified the option to base an algorithm on the energy of the aircraft at specific points during the approach, generating two more algorithms. Each of these algorithms is discussed in more detail below.

1) Criteria by the FSF ALARP Task Force

The criteria for a stable approach defined by the Flight Safety Foundation Task Force [5] are the industry standard and have been incorporated in the airline's standard operating procedures with the following modification: (1) the airplane should be at approach speed, deviations of +10 knots to -5 knots are acceptable if the airspeed is trending toward approach speed; and (2) sink rate is no greater than 1,000 fpm. Earlier studies have identified that pilots' assessment of an unstable approach that warrants a go-around match these criteria, but use limits that are much less stringent [3]. The criteria are judged from 500 ft AAL (above aerodrome level) to flare (50 ft AAL).

2) 10 NM Limits

The airline's standard operating procedures define a separate set of criteria to be met at 10 NM track distance to touchdown: "It is good operating practice to be at 3,000 feet above field elevation at 10 NM track miles from touchdown with flaps 1 and speed maximum 210 kts. These criteria are to be considered limits and not targets. Use speed brakes and consider lowering the landing gear early when deviating above the profile." These precautionary limits should reduce the chance of an unstable approach, and are considered to be a possible algorithm to detect unstable approaches.

3) Aircraft Energy

According to [17], "approximately 70 % of rushed and unstable approaches involve an incorrect management of the aircraft energy level, resulting in an excess or deficit of energy." Therefore it of interest to research what the energy can have for predictive value on unstable approaches. This is investigated for both stabilization gates. The total energy of the aircraft (E_{tot}) can be calculated with the kinetic and potential energy of the aircraft at any given moment. The kinetic energy of the aircraft depends on the aircraft mass and speed:

$$E_{kin} = 0.5 \cdot m \cdot v^2 \qquad (1)$$

Where E_{kin} is the kinetic energy [J], m = mass [kg] and v = speed [m/s].

The potential energy depends on the weight of the aircraft and the height above the earth's surface:

$$E_{pot} = m \cdot g \cdot h \qquad (2)$$

Where E_{pot} is the potential energy [J], m = mass [kg], g = gravitational constant [m/s²], and h = height [m].

To enable comparison of energy levels of different flights, it is necessary to define a factor for the energy level that is independent of aircraft mass:

$$EF = (E_{pot} + E_{kin} - E_{opt})/E_{max} \qquad (3)$$

Where EF is the mass independent energy factor, E_{pot} and E_{kin} as defined before in (1), (2); E_{opt} = Optimal energy, i.e. the energy when flying on the glide path and with the correct approach speed (V_{app}); and E_{max} = Maximum energy, i.e. the energy when the aircraft is one dot above the glide path and with a speed of 10 knots above the approach speed, all in [J]. An EF of 0 indicates an optimal energy content. An EF of 1 indicates a flight path that is one dot above glide path and an airspeed of V_{app} +10 kts, which is a barely permissible deviation. There is not a fixed value for the barely permissible downward deviation of EF. For every approach therefore an E_{min} is calculated, based on the minimum value for the speed V_{app}-5 knots and one dot under the glide path.

4) Energy at 10 NM Track Distance

To see if unstable approaches can be predicted with the energy of the aircraft even earlier in time the energy level at 10 NM track distance before touchdown will be looked at. Only maximum limits for the energy are described in the airline's operating procedures. The maximum permissible energy for the 10 NM point is when the aircraft is at 3000 ft AAL and has an airspeed of 210 kts. To be able to calculate the EF the optimal energy level of the aircraft needs to be defined. The defined optimal energy level at the 10 NM point will be when the aircraft is on the glide path of 2.65 degrees (one dot below the regular three degree flight path) and flies an airspeed of 200 kts. The airspeed of 200 kts is chosen because the maximum airspeed of 210 kts is seen as the V_{app} +10 kts speed which makes 200 kts the replacement for the V_{app}. This is done so the EF at 10 NM track distance can be calculated in the same way as the EF at 500 ft and 1000 ft AAL (see above).

D. Research Questions

Based on the need for an algorithm to automatically identify genuinely unstable approaches and the potential algorithms that have been derived from the literature, the following research questions are formulated:

- How do we differentiate genuinely stable from genuinely unstable approaches?
- Which algorithm most closely correlates with the categorization by flight data analysts of genuinely stable and unstable approaches?
- What conclusions can we draw from the discrepancy between the current industry standard for unstable approaches and the criteria for genuinely unstable approaches?

These questions will be answered in the remainder of this paper.

II. METHOD

A. Identification of genuine unstable approaches

To identify genuine unstable approaches, a set of 30 flights is extracted from the flight data management system that are not chosen randomly, but rather have been selected to be close to the border between genuinely stable and unstable. These flights have not previously been evaluated by flight data analysts (FDA's). The flights are reviewed independently by four FDA's, consisting of two safety engineers and two investigator pilots. For each flight the FDAs are asked to categorize the approaches as genuinely unstable or not, and to indicate the reasons for their decision.

To calculate the level of agreement of their decisions, Cohen's Kappa is used. Cohen's Kappa is a statistical measure of agreement between two parties. The number of full agreement in decisions is tested against the number of decisions on which two parties did not agree. When the outcome of the Cohen's Kappa coefficient (κ) is below 0.4 it can be considered as a poor level of agreement. Between 0.4 and 0.6 it is stated to be moderate and between 0.6 and 0.8 the result is a good level of agreement. Above 0.8 it defines an excellent agreement level [18]. For more than two parties, pairwise calculation of Cohen's Kappa is often used.

B. Algorithm Design

To set up a good working algorithm it was important to know which parameters were monitored when flights were reviewed by the FDAs. For this reason interviews, observations and an assessment list were used. It was specified to the FDAs that they should base their decisions only on what occurs below 500 ft AAL With the use of the information from the literature review and the assessment list of the FDAs the algorithms could be optimized to match the assessment of the FDA's.

To find the best parameter setting for each algorithm in an iterative manner, it was analyzed which boundaries led to the best agreement with the combined decision of the FDA's for the data set of 30 flights. In case of multiple criteria (i.e. the algorithms based on the criteria by the FSF ALARP Task Force and 10 NM out) it was identified for each criterion whether (1) the approach is stable, (2) the approach is unstable due to another parameter, (3) the approach is unstable due to this and another parameter, or (4) the approach is unstable due to solely this parameter. By first setting the limits through (1), (2) and (4) and then adjusting through (3) an optimal algorithm is defined.

C. Algorithm Validation

The optimal algorithms are validated against a second set of 30 flights (chosen independently of the first set) to prove their effectiveness.

D. Integration into the flight data management system

Once an algorithm has been selected that closely matches the current evaluation of the flight data analysts, this can be programmed into the flight data management system. The current flight data management system at the airline is Aerobytes (www.aerobytes.co.uk). This program holds all the available flight data of the executed flights. Aerobytes has the option to program algorithms into the software itself. Such an algorithm is called a state. When the conditions of the state are met, the

state is "detected" in a flight. If these conditions are the boundaries of the unstable approaches algorithm, unstable approaches can be detected by Aerobtyes.

III. RESULTS

A. Identification of genuine unstable approaches

Thirty flights have been chosen from the flight data management system. 29 of these do not meet the FSF ALRP Task Force criteria for stable approaches. The results of the FDA evaluation are given in Table 1. As can be seen in there is unanimous agreement on twenty of the flights.

TABLE 1: RESULTS FOR 30 FLIGHTS PER FDA

Flight Nr.	FDA 1	FDA 2	FDA 3	FDA 4
1	No	No	No	No
2	No	Yes	No	No
3	No	No	No	No
4	No	No	No	No
5	Yes	Yes	No	Yes
6	Yes	Yes	No	No
7	No	No	No	No
8	No	Yes	No	No
9	No	No	No	No
10	No	No	No	Yes
11	Yes	Yes	Yes	Yes
12	Yes	Yes	Yes	Yes
13	Yes	Yes	Yes	Yes
14	Yes	Yes	Yes	Yes
15	No	Yes	No	No
16	No	No	Yes	No
17	Yes	Yes	Yes	Yes
18	No	No	No	No
19	Yes	Yes	Yes	Yes
20	Yes	Yes	Yes	Yes
21	No	No	No	No
22	No	No	No	Yes
23	Yes	No	No	No
24	No	No	No	No
25	No	No	No	No
26	No	No	No	No
27	No	No	No	No
28	Yes	Yes	Yes	No
29	Yes	Yes	Yes	Yes
30	Yes	Yes	Yes	Yes

A flight will be considered genuinely unstable or stable if three or more FDA's have indicated it as such. One flight (flight 6) is eliminated from the data set because there was no clear decision.

Cohen's Kappa has been calculated for these results. As can be seen in Table 2 there is moderate to good agreement between the FDA's.

TABLE 2: COHEN'S KAPPA FOR 30 UNSTABLE APPROACHES

		FDA			
		1	2	3	4
FDA	1		0.733	0.724	0.658
	2			0.600	0.533
	3				0.648
	4				

B. Algorithm design

During the interviews with, and observations of, the FDA's it became clear that they focused on the following parameters when assessing a genuinely unstable approach: Rate of Descent, Roll, Airspeed, Thrust setting, Flap setting, and Gear position. These were used (where possible) to optimize the algorithms that were introduced in section I.

1) Criteria by the FSF ALARP Task Force

The FSF ALARP Task Force has defined criteria for a stable approach incorporating: vertical and horizontal flight path; airspeed; descent rate; and thrust, gear, and flap settings. These have been incorporated in the airline's standard operating procedures (SOP).

a) Rate of Descent

The rate of descent is defined in the SOP and may not exceed the limit of 1000 fpm. However, the assessors do not consistently consider an approach as unstable when this limit is exceeded. It seems all assessors do not adopt a set limit for this, but keep this in mind. From interviews it was noticed that the border of 1100 fpm seems to have a decisive function. The extent (in steps of 50 fpm) to which the limit is exceeded is compared with the duration of the breach (in seconds).

b) Roll

There was a discrepancy in decision making depending on whether a circling approach was executed or not. A circling approach is an approach where the aircraft flies parallel to the runway shortly before the landing. In some approaches this leads to a late turn before landing. The SOP prescribes that during a circling approach wings should be levelled at 300 ft AAL. The discrepancy in decision making led to two analyses for the bank angle: under 500ft and 300 ft and lower. The time the bank angle exceeds 5 degrees, 8 degrees and 10 degrees is taken as variable.

c) Airspeed

The FDAs do not draw a hard line concerning the airspeed limits. The time the airspeed exceeds the approach speed plus 10 kts is measured. Also, the time the aircraft exceeds the approach speed plus 12 kts, plus 15 kts and plus 20 kts is measured so four different points are used. Solely an exceedance of this single parameter was sufficient for FDAs to consider the approach to be unstable.

d) Thrust setting

The power setting is of concern for the assessors and influences the decision making process. The reason for this is safety related and can be explained when imaging a landing with

a low power setting. Before the engines are at a setting at which a Go Around can be executed, valuable time and altitude is lost. The SOP prescribes the thrust setting as "is appropriate", but does not provide an exact value. To identify a possible border, the thrust setting is examined and plotted against the time the thrust setting is below 50%, 45%, 40% and 30%. These values represent the percentage of rotational speed of the shaft of the engine (N1). Since both engines are (or should be) equal, only the left one is used for the analysis. High power settings are plotted in the same way as described above for thrust settings above 70%, 75%, 80% and 85%. The latter however was in none of the approaches reason to label it as unstable. Low power settings were in several approaches reason to designate the approach as unstable. Often this is found in combination with other violations of the criteria. This is explained by the effect of a higher thrust setting. If the aircraft has too much speed or a high rate of descent, a higher power setting would counterwork the desired state. Since the power setting can easily be adjusted it is often used to bring other parameters back within limits. Therefore, when other parameters are too high, the thrust setting is often found to be too low.

e) Flap setting

If an aircraft is not configured to the landing flap setting before 500 ft AAL, the assessors conclude an unstable approach immediately. Not having the landing flaps selected indicates an incomplete landing checklist because it is the last item to check. This often indicates the crew cannot fully concentrate on the landing. Of all approaches the landing flap selection is analysed and the associated height is recorded.

f) Gear position

Throughout interviews it was noticed that the gear position is a decisive parameter in the decision of an assessor. Due to the time it takes to extend the gear from the start position it is required to ensure the gear is extended before 500 ft AAL. Nevertheless, this extension has the function to increase drag in the approach and indicates the crew progress towards the landing. It is assumed that the retracted position of the gear directly leads to an unstable approaches. Unfortunately, this assumption was not testable since no approaches were found where the gear was not extended at 500 ft AAL .

g) Final result for the adapted ALARP criteria

As a result of the above calculations, the criteria according to the FSF ALARP Task Force can be modified. It is found that a flight is genuinely unstable under the following conditions:

- The rate of descent exceeds 1100 fpm for 4 seconds or more.
- The rate of descent exceeds 1050 fpm for 6 seconds or more.
- The rate of descent exceeds 1000 fpm for 8 seconds or more.
- The bank angle exceeds 10 degrees for 2 seconds or more.
- The bank angle exceeds 8 degrees for 4 seconds or more.
- The bank angle exceeds 5 degrees for 10 seconds or more.
- The bank angle exceeds 10 degrees for 1 second or more at or under 300 ft AAL.
- The bank angle exceeds 8 degrees for 2 seconds or more at or under 300 ft AAL.
- The bank angle exceeds 5 degrees for 3 seconds or more at or under 300 ft AAL.

- The airspeed exceeds the approach speed plus 10 for 12 seconds or more.
- The airspeed exceeds the approach speed plus 15 for 2.5 seconds or more.
- The thrust setting for engine 1 is below 37% for 1 second or more in the range of 300 ft AAL to 50 ft AAL.
- The thrust setting for engine 1 is below 35% for 8 seconds or more in the range of 500 ft AAL to 50 ft AAL.
- The thrust setting for engine 2 is below 37% for 1 second or more in the range of 300 ft AAL to 50 ft AAL.
- The thrust setting for engine 2 is below 35% for 8 seconds or more in the range of 500 ft AAL to 50 ft AAL.
- The landing flaps are not selected.
- The landing gear is not extended.

These limits have been defined based on the 29 initially chosen flights. Therefore, the correlation with the FDA evaluation of these is excellent: Cohen's Kappe (κ) = 0.925. One flight is identified as genuinely unstable by the algorithm while the FDA's agree on its stability. All other 28 flights are coded correctly.

2) 10 NM Limits

The parameter values at 10 NM track distance from touchdown have been similarly analysed for altitude (> 3200 ft AAL or < 1600 ft AAL), airspeed (>230 kts or < 170), and flaps (first setting not selected). A comparison with the judgement of the FDAs shows a good correlation: Cohen's Kappa κ = 0.640. 5 out of 29 flights are not coded correctly.

3) Aircraft Energy

The EF at 1000 ft is calculated using EF = 0.8 as the high limit and EF = -0.4 as the lower limit. There is a moderate correlation (κ = 0.513), although 7 out of the 29 flights are not identified successfully.

The EF at 500 ft is calculated using EF = 0.5 as the high limit (no low limit). There is a limited correlation with the FDA's (κ = 0.434). Eight flights are not coded correctly.

4) Energy at 10 NM Track Distance

The EF is calculated using the EF = 0.5 as the high limit and -1.9 as the lower limit. There is moderate correlation with the FDA's (κ = 0.496). Seven flights are not coded correctly.

C. Algorithm Validation

Validation of all the algorithm will be attempted on a new data set of 30 flights.

1) Identification of genuine unstable approaches

The approaches in the second dataset are selected by the algorithm results from the first set of approaches. The selected approaches for the validation were chosen to be close to the conditions of the statements from the optimal algorithms. This led to a set of approaches of which the clarity between stable and unstable was even lower than in the first assessment list. For each flight the FDAs are asked to categorize the approaches as genuinely unstable or not, as shown in Table 3.

TABLE 3: FDA DECISION ON SET OF 30 VALIDATION FLIGHTS

Flight Nr.	FDA 1	FDA 2	FDA 4	Decision
1	Yes	Yes	Yes	Yes
2	No	Yes	No	No
3	No	Yes	No	No
4	No	Yes	No	No
5	No	Yes	No	No
6	Yes	Yes	Yes	Yes
7	Yes	Yes	Yes	Yes
8	No	No	No	No
9	No	No	No	No
10	No	Yes	Yes	Yes
11	No	No	No	No
12	Yes	Yes	No	Yes
13	Yes	Yes	Yes	Yes
14	No	Yes	No	No
15	No	Yes	Yes	Yes
16	Yes	Yes	Yes	Yes
17	Yes	Yes	Yes	Yes
18	Yes	No	No	No
19	No	No	No	No
20	Yes	Yes	No	Yes
21	Yes	Yes	Yes	Yes
22	No	Yes	No	No
23	No	Yes	No	No
24	Yes	Yes	No	Yes
25	Yes	Yes	Yes	Yes
26	Yes	Yes	Yes	Yes
27	No	Yes	No	No
28	Yes	No	Yes	Yes
29	Yes	No	Yes	Yes
30	Yes	Yes	Yes	Yes

Because the approaches are selected close to the found algorithm statements, it is expected that the FDA's find the second assessment list harder to analyze than the first list, and a decrease in agreement is expected between the FDAs. The correlation between the FDA judgments are shown in Table 4.

TABLE 4: COHEN'S KAPPA FOR 30 VALIDATION APPROACHES

		FDA		
		1	*2*	*3*
FDA	*1*		0.101	0.602
	2			0.163
	3			

As can be seen the level of agreement is quite low, indicating the borderline characteristics of the validation flights between genuinely stable and genuinely unstable. It is therefore natural that the algorithms will also show a lower correlation (Cohen's Kappa) in the validation than reported above.

2) Validation of the adapted ALARP criteria

The adapted ALARP criteria algorithm categorizes 23 of the 30 lights correctly as genuinely stable or unstable. 7 flights are categorized as unstable where in fact the majority of the FDA's considered these as stable (type I error). The type I errors are seen as less significant compared to type II errors because the error is on the side of safety. Cohen's Kappa shows moderate correlation ($\kappa = 0.553$).

3) 10 NM Limits

The optimal algorithm based on the limits from the SOP at 10 NM track distance shows poor level of agreement ($\kappa = 0.186$). The algorithm classifies twelve approaches incorrectly.

4) Aircraft Energy

The algorithm for the Energy Factor at 1000 ft AAL has a maintained a moderate level of agreement with the decisions of the FDAs on the validation list ($\kappa = 0.521$). The total amount of mistakes made by the algorithm is the same: seven.

The algorithm based on the EF at 500 ft AAL resulted in a poorly scoring algorithm: $\kappa = 0.273$. Eleven flights were not classified correctly.

5) Energy Factor at 10 NM Track Distance

This algorithm performs very poorly ($\kappa = 0.032$). Nearly half of the flights (14 out of 30) are not coded correctly.

IV. DISCUSSION & CONCLUSION

In this paper we have presented the results of a study into the automatic identification of genuinely unstable approaches. This research was inspired by the need to alleviate the manual task involved in identifying these genuinely unstable approaches. Using the industry standard (the FSF ALARP Task Force criteria, [5]) was not appropriate as it flagged too many flights that were only marginally unstable and that did not warrant further discussions. It is not practical to discuss about 4% of the flights that violate the criteria for a stable approach, particularly because no reference is made to the extent or duration of the violation, or the combination of criteria that are breached. Four algorithms (adapted from ALARP criteria, configuration at 10 NM from touchdown, Aircraft energy at 500 and 1000 feet, and energy 10 NM out) that were derived from the literature were optimized based on a set of thirty flights, and then validated on a separate set of flights. The assessment as to whether a flight was genuinely unstable or not was based on the majority judgments of flight data analysts.

Of the four algorithms the algorithm adapted from ALARP criteria performed best with a moderate level of agreement on the validation set. This is a result of the wide range of detection possibilities: seventeen time-dependent limits for descent rate, bank angle, airspeed, thrust settings flaps and landing gear over a time range from 500 ft AAL to flare. The limits that have been defined are significantly less stringent than those recommended by the FSF ALARP Task Force, and mirror those that pilots use in their assessment of an unstable approach and the need to initiate a go-around [3].

Of all 59 flights that were analyzed by the algorithm as well as the FDA's, only in 8 cases was there a mismatch. In all cases the error was of type I. This type of error is preferred over type II errors since false alarms are preferred over missed calls. Note that the 59 flights were not randomly selected and therefore not a representative selection of approaches: all but one constituted an unstable approach according to the original more stringent FSF ALARP Task Force criteria.

The algorithms based on the EFs and the 10 NM limits are calculated at only one moment in time. The energy levels for the aircraft generally fluctuate over a short range of time. This is not of benefit for the selection of the optimal limits because a few moments earlier or later the EF can be so different. Additionally, the EF does not take all the decisive parameters from the FDAs into account; the flap setting and the moment the landing gear is extended are eliminated in the EF calculation. For the two algorithms based on the 10 NM point a larger amount of time to stabilize the approach is available. Therefore less correlation with the FDA assessment was possible.

The algorithm adapted from ALARP criteria has since been implemented in the Flight Data Management System Aerobytes at the airline. It alerts FDA's automatically about genuinely unstable approaches. This is the case in 2-3% of the approaches (versus > 10% using the original criteria). Additionally, 26,044 flights have been analyzed, to identify the frequency of genuinely unstable flights over the period May 2010 – April 2013 for different aircraft types and destinations for the airline. The (confidential) analysis shows a very large variation between different airports, with some destinations showing percentages far above 50%. This implies that local circumstances (late turns to final, terrain, wind shear) strongly dictate whether a flight will be genuinely unstable or not.

As was shown particularly with the validation set, the perception of genuinely unstable approaches differs from FDA to FDA. This discrepancy can possible be ascribed to the difference in job function between safety engineers and pilots. The validation list supports this presumption, pilots seem to classify approaches as stable more often than safety engineers. This discrepancy can be explained by a higher empathy with the pilots flying the approach and knowing how they themselves would act in a similar situation. Safety engineers on the other hand are more likely to be strict because they want to maintain the safety in the flight operations.

The results of this study suggests that the FSF ALARP Task Force criteria for unstable approaches may be too strict, particularly if they are interpreted as a trigger for a go-around [4], [5]. Local circumstances often make an unstable approach inevitable, and pilots consider a flight genuinely unstable only if certain flight criteria are breached to some extent and for a number of seconds. This finding matches the earlier results of Wischmeyer [16], who suggests that an unstable approach (using the traditional definition) cannot be considered a safety risk. The low number of go-arounds (3%) that are initiated in an unstable approach [3] may well be the result of the rational deliberation between the risk of a go-around versus continuing the approach. In fact, if all approaches that breach the FSF ALARP Task Force criteria initiate a go-around, this will result in a dramatic lowering of airport capacity [11]. It will quite possibly also introduce new safety risks, despite the fact that go-arounds are considered "a normal phase of flight and the operational risk associated with this phase should be comparable to those related to other phases" [6]

We suggest that a new definition of a genuinely unstable approach be adopted, that specifies the extent and duration that limits may be breached for the deviation from the glide scope, levelling of the wings, rate of descent, thrust setting, track, airspeed, gear and flaps. This less stringent definition will allow a focused effort on go-around initiation under these more risky conditions. Further research will identify whether this more limited set of genuinely unstable approaches in fact does pose a safety risk.

REFERENCES

[1] B. Smith, M; Curtis, "Why are Go-Around Policies Ineffective ?," in *Go-around Safety Forum, Brussels*, 2013.

[2] M. Kroepl and G. Burton, "STEADES High-level analysis," in *Go-around Safety Forum Brussels*, 2013.

[3] J. M. Smith, D. W. Jamieson, and W. F. Curtis, "Why are go-around policies ineffective ? The psychology of decision making during unstable approaches," in *65th Annual FSF International Air Safety Seminar Santiago, Chile*, 2012.

[4] Airbus Customer Services, "Flight Operations Briefing Notes Approach and Landing (FLT_OPS – GEN – SEQ 01 – REV 03)," Blagnac, France, 2004.

[5] Flight Safety Foundation Editorial Staff, "FSF ALAR Briefing Note 7.1: Stabilized Approach," Alexandria, VA, 2009.

[6] Flight Safety Foundation Editorial Staff, "Findings and Conclusions," in *Go-around Safety Forum Brussels*, 2013.

[7] B. De Courville, "Go-around Decision Making," in *Go-around Safety Forum, Brussels*, 2013.

[8] J. Wenderich, "Learning about unstabilized approaches through animations," in *Go-around Safety Forum Brussels*, 2013, no. May.

[9] M. Kroepl and G. Burton, "STEADES In-depth analysis," in *Go-around Safety Forum Brussels*, 2013.

[10] M. Heiligers, T. Holten, and M. Mulder, "Flight Mechanical Evaluation of Approaches," *J. Aircr.*, 2011.

[11] A. Gammicchia, "Pilot decision-making," in *Go-around Safety Forum Brussels*, 2013.

[12] R. Khatwa, R. & Helmreich, "Analysis of Critical Factors During Approach and Landing in Accidents and Normal Flight – Data Acquisition and Analysis Working Group Final Report.," *Flight Saf. Dig.*, 1999.

[13] Australian Safety Transport Bureau, "Runway excursions Part 1: A worldwide review of commercial jet aircraft runway excursions," 2008.

[14] W. Rosenkrans, "Overrun Breakdown," *AEROSAFETYWORLD*, no. November, pp. 8–11, 2012.

[15] E. Pooley, "GO AROUND ACCIDENT & INCIDENT REPORT REVIEW," in *Go-around Safety Forum Brussels*, 2013, no. June.

[16] E. Wischmeyer, "The Myth of the Unstable Approach," in *International Society of Air Safety Investigators*, 2004.

[17] Airbus Customer Services, "Flight Operations Briefing Notes FLT_OPS – APPR – SEQ03 – REV02," Blagnac, France, 2005.

[18] C. Bérard, C. Payan, I. Hodgkinson, and J. Fermanian, "A motor function measure scale for neuromuscular diseases. Construction and validation study," *Neuromuscul. Disord.*, vol. 15, no. 7, pp. 463–470, Jul. 2005.