Go-Around Criteria Refinement for Transport Category Aircraft

Angela M. Campbell[*]
Federal Aviation Administration
Atlantic City, NJ, USA

Peter M. T. Zaal[†]
Metis Technology Solutions, Inc.
NASA Ames Research Center
Moffett Field, CA, USA

Somil R. Shah[‡]
Federal Aviation Administration
Atlantic City, NJ, USA

Jeffery A. Schroeder[§]
Federal Aviation Administration
Moffett Field, CA, USA

Presently, airline pilots are trained to go around if, when lower than 500 ft above the ground, they are outside of a handful of parameters such as airspeed, position, and rate-of-descent. At times, pilots do not comply with these criteria, perhaps owing to their conservative nature or complexity. This paper examines potential refinements to the continue-to-land decision from the combined results of three flight simulator experiments. Potential refinements include simplifying the number of parameters and lowering the altitude at which pilots make the decision. First, refinements were developed by evaluating pilots' touchdown performance and qualitative data in a variety of starting and environmental conditions. Second, 30 of those pilots evaluated the refinements under several induced instabilities during the approach. The results showed little difference in touchdown performance when lowering the decision altitude from 500 ft to 300 ft; however, significant differences arose when the decision altitude was lowered further to 100 ft. The proposed new criteria include assessments of deviations in airspeed and

[*]Aerospace Engineer, William J. Hughes Technical Center, Atlantic City International Airport, NJ, 08405; angela.campbell@faa.gov. Member.

[†]Principal Aerospace Engineer, SimLabs, NASA Ames Research Center, Moffett Field, CA, 94035; peter.m.t.zaal@nasa.gov. Associate Fellow AIAA.

[‡]Aerospace Engineer, William J. Hughes Technical Center, Atlantic City International Airport, NJ, 08405; somil.shah@faa.gov. Member.

[§]Chief Scientific and Technical Advisor for Flight Simulation Systems, Moffett Field, CA, 94035; jeffery.schroeder@faa.gov. Associate Fellow AIAA.

The findings and conclusions in this paper are those of the author(s) and do not necessarily represent the views of the FAA.

This paper is related to work published and presented at the 2018 AIAA Aviation Forum (Atlanta, GA; 25-29 June 2018; AIAA 2018-319) and the 2019 AIAA Aviation Forum (Dallas, TX; 17-21 June 2019; AIAA 2019-2993 and AIAA 2019-3610)

position, no rate-of-descent audio warning, and having an appropriate engine setting at 1,000 ft, 500 ft, and 300 ft height above threshold . Additionally, a recommendation is made that if the proposed criteria are not met at the 1,000 ft or 500 ft height above threshold the pilots may make corrections and continue the approach, however, if the criteria are not met at 300 ft, then a go around should be performed.

Nomenclature

h = height above ground, ft

$\dot{h}$ = rate-of-descent, ft/min, ft/s

V = airspeed, kts

x = longitudinal distance, ft

y = lateral distance, ft

Δ = deviation

μ = mean, –

σ = standard deviation, –

Subscripts

ga = go-around

ref = reference

tgt = target

td = touchdown

I. Introduction

Stabilized approach criteria for transport category aircraft have fallen under scrutiny due to low compliance with established policies. Most airlines have defined stabilized approach criteria in their standard operating procedures. Because the criteria are set by each operator, the criteria and specified altitudes (or gates) vary across the industry as shown in Table 1. Typically, airline procedures state that if the pilots determine that the approach is unstable based on a predefined set of criteria, they should conduct a go-around, or execute a missed approach. However, the collective industry performance of complying with go-around policies is extremely poor, with only approximately 3% of unstable approaches resulting in a go-around [1].

Studies by aviation industry safety groups have suggested that the go-around noncompliance rate could be a significant safety hazard for commercial aviation given that approach and landing are the most common phases of flight for aviation accidents, accounting annually for approximately 65% of all accidents [2]. A Flight Safety Foundation (FSF) study of 16 years of runway excursions found that 83% of them could have been avoided by a decision to go around; thus, 54% of all accidents could potentially have been avoided by the execution of a missed approach [1].

Improving the go-around compliance rate holds significant potential in reducing approach and landing accidents; consequently, organizations such as the Commercial Aviation Safety Team (CAST) [3], the International Air Transport Association (IATA) [2], and the FSF [1] have investigated the root causes of go-around noncompliance and have made recommendations for improving compliance rates. Studies by these organizations have revealed many reasons for noncompliance by flight crews related to pilot judgment and company policies, including little-to-no consequence for not following policies, lack of management awareness of the compliance rate, as well as pilot fatigue and situation awareness. However, one of the largest factors is that pilots view the current stabilized approach criteria as too complex or restrictive for the modern operational environment. For example, there are as many as eight parameters that can be found in stabilized approach criteria which is a large number of parameters for a pilot to track in a dynamic operational environment. Additionally, in some definitions of stabilized approach criteria, the tolerances on a parameter can be small. For instance, criteria can have ranges as small as 10 knots for airspeed which can be difficult to meet in gusty conditions.

Table 1 Typical stabilized approach criteria.

	Parameter	Threshold
1	Stabilization height	1,000 ft IMC, 500 ft VMC
2	Localizer deviation	± 1/2-1 dot
3	Glideslope deviation	± 1 dot
4	V_{ref} deviation	V_{ref}-5 – V_{ref}+(10-20) kts
5	Rate of descent	1,000 ft/min
6	Power on approach	Above idle
		Appropriate with airspeed,
		approach, and condition
7	Aircraft configuration	Flaps in landing conf.
		Landing gear deployed
		Spoilers armed
8	Bank angle	Wings level 300-500 ft
		Less than 25 deg

Establishing an industry standard for missed approach criteria is challenging because of the large number of variables that influence approach and landing risk, such as aircraft state, human factors, and environmental conditions. Additionally, the criteria must be able to mitigate the approach and landing risk without being so restrictive that they cannot be realistically implemented in today's operational environment. Recently, several airlines and safety organizations, including the FSF, have established revised stabilized approach criteria based on their approach and landing risk assessments. One example of a revised set of criteria is the guidelines published by the FSF in a 2017 report. These guidelines suggest that the go-around decision height could be lowered to 300 ft above ground level (AGL). The 500-ft or 1,000-ft gates that are normally used by most airlines today could be re-classified as stabilized approach gates where the stability of the approach could be checked but the approach may continue if deviations from the criteria are able be corrected by the 300-ft gate [1].

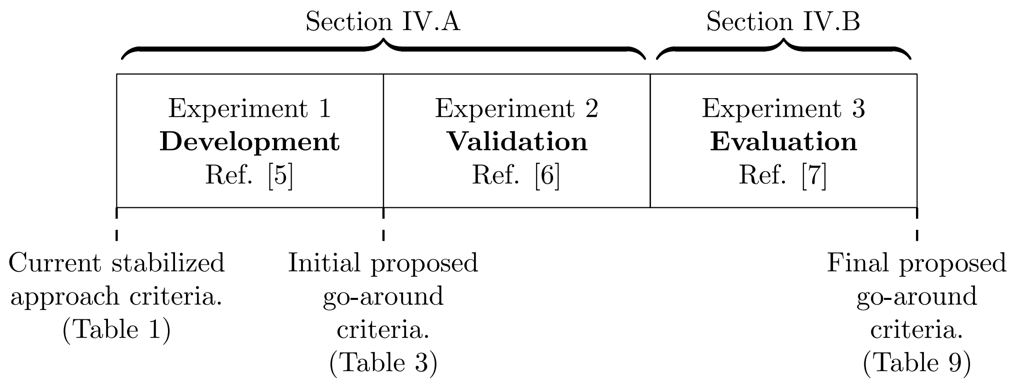

Fig. 1 Study overview and paper structure.

In addition to studies conducted by IATA and the FSF, the Federal Aviation Administration (FAA) has established a research program to evaluate the current stabilized approach criteria guidance such as Advisory Circular (AC) 91-79A [4]. The overarching goal of this research is to examine potential refinements to current go-around criteria and validate recommendations from organizations such as the FSF [5–7]. This paper presents the results of the combined data of three experiments conducted as part of this research that developed, validated, and evaluated a new set of universal, simplified go-around criteria (Fig. 1). As such, only key results and findings from each experiment will be included in this paper; additional details can be found in [5–7]. Section II provides an overview of the simulators and pilot participants used in the experiments. Section III provides an overview of the design and methodology used for each experiment as well as its key findings. The results are presented and discussed in Section IV and Section V, respectively. Finally, conclusions made from the overall study are provided in Section VI.

II. Flight Simulators and Pilot Participants

Experiment 1 used three Level D full flight simulators: a Boeing 747-400 simulator at NASA Ames Research Center in Mountain View, California [8], and Boeing 737-800 and Airbus A330-200 simulators at the FAA Mike Monroney Aeronautical Center in Oklahoma City, Oklahoma. Experiments 2 and 3 utilized only the latter two simulators. The three aircraft types tested provided the ability to compare results among both narrow-body and wide-body aircraft, as well as among the Boeing and Airbus flight decks. This approach adds confidence to the widespread applicability of the results to the fleet.

All simulators were from the same manufacturer and used in their standard configurations. Differences between simulators existed because of the different years of initial operation and the different aircraft types simulated. Care was taken to make all basic aircraft and environmental settings as similar as possible between simulators. For example, cockpit radio-altitude and warning call-outs, turbulence intensity, and runway and radio-navigation-aid geometries were equalized to provide pilots with similar basic cues across simulators. The out-the-window visual database was different between the A330/B737 and B747 simulators because of the different visual system manufacturers; however, basic features in the visual scene were equalized. Motion cues were provided in each simulator using the standard motion logic settings.

All approaches and landings were flown without any automation engaged; the autopilot, autothrottle, autobrakes, as well as the flight director were turned off for all approaches in all simulators. The primary flight display (PFD) depicted the conventional localizer (LOC) and glideslope (GS) error indicators. For each experimental condition, speed deviations were provided as deviations from the target approach speed, V_{tgt}, of the specific aircraft. The target approach speed is the reference speed, V_{ref}, corrected for the wind condition (in this case a 10-kt tail wind). The target speeds for the different aircraft types utilized were 141, 148, and 153 knots indicated airspeed (KIAS) for the A330, B737, and B747, respectively. Fig. 2 shows the PFD in the B737-800 simulator, with the corresponding out-the-window visual provided in Fig. 3. In these figures, the aircraft is at the 500-ft gate, 1.5 dots above the glideslope, 1.5 dots to the right of the localizer, descending at 1,500 ft/min, and at an initial airspeed of V_{tgt} + 20 KIAS. Note that the aircraft is approximately lined up with the taxiway to the right of Runway (RWY) 28R at San Francisco International Airport (SFO).

Fig. 2 B737-800 PFD.

Fig. 3 Out-the-window view.

Six crews comprised of a captain and a first officer from the same airline participated in each simulator for all experiments, resulting in a total of 18 crews (or 36 pilots) for Experiment 1 and a total of 12 crews (or 24 pilots) for Experiments 2 and 3. Crews from six different airlines participated. Most pilots were experienced with ratings on multiple aircraft types. One Boeing 737 pilot was retired at the time of the experiment. All other A330 and B737 pilots were current and qualified as captain or first officer in a Part 121 carrier. It was a requirement for Boeing 747 pilots to be flying for a Part 121 carrier, and current and qualified as captain or first officer in the B747-400 in the last 36 months prior to the experiments. All pilots gave written consent for their participation and received compensation.

The goal of using the full flight simulators and current and qualified pilots was to create a realistic response to what would occur in real operations. However, the study does have some limitations which should be noted:

1) The intent was to recruit pilots that are representative of pilots flying in the United States; however, due to the nature of the recruiting process and available pilots in the recruiting pool, some biases may be present in the sample.
2) The study was conducted using narrow-body and wide-body aircraft simulators, and included both Boeing and Airbus aircraft which represent a large percentage of the commercial transport fleet. However, caution is encouraged when extrapolating the results of this study to regional jets, turboprops, or other aircraft types.
3) In the simulation environment, the crew is not exposed to operational pressures such as passengers and schedule. Therefore, a conservative interpretation of the results should be used.
4) All approaches were flown manually in order to assess the ability of the pilots to correct an unstable approach and to avoid incapability issues between the simulator and some scenarios. Therefore, the workload associated with monitoring the automation or correcting an unstable approach while automation is engaged were not covered in this study.

III. Experimental Method

This section provides an overview of each experiment as depicted in Fig. 1. Additional details and results for experiments 1, 2, and 3 can be found in [5], [6], and [7], respectively.

A. Experiments 1 and 2 - Development and Validation of Stable Approach Criteria

1. Approach and Forced Landing Task

The basic premise of Experiment 1 was to correlate touchdown performance (dependent measures) with various approach states and environmental conditions (independent variables) by requiring pilots to land the aircraft within a defined touchdown zone under a variety of initial-approach conditions (starting conditions or scenarios). In Experiment 1, touchdown performance was evaluated for various approach states only. These results were then validated in Experiment 2 using different approach states and environmental variables such as changes in wind and visibility.

Pilots participating in experiments 1 and 2 of the study were required to always land the aircraft (no go-arounds were allowed), even from conditions considered to be unstable by their airline and regardless of whether or not they personally felt that a go-around should be conducted. The reason for this decision in the experiment design was to remove the go-around decision-making process and to obtain an objective assessment of the likelihood of an abnormal landing under various approach states. The expectation was that, under certain scenarios, pilots would do the following: (a) land outside of the specified touchdown zone; (b) have an excessive ground speed at touchdown; or (c) have an excessive sink rate at touchdown. Under these conditions, the risk of an accident or incident would be elevated in reality. Using results from the first two experiments, a determination can be made regarding the limits for which a pilot can land safely with an acceptable level of risk.

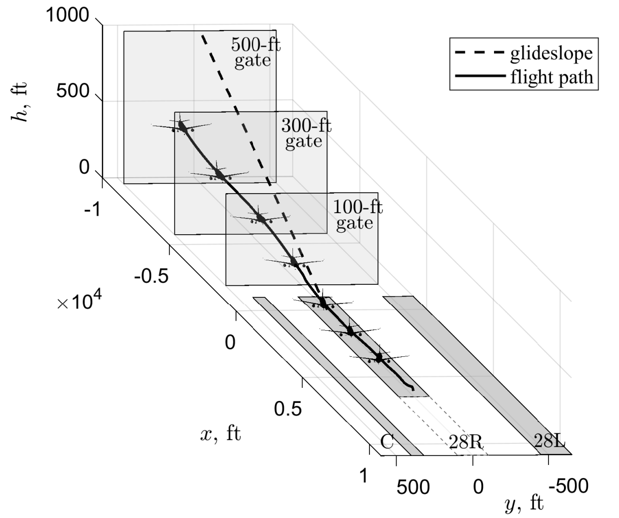

Fig. 4 Mean B737 trajectory with lateral instability for Experiment 1.

Aircraft: B747-400, B737-800, and A330-200
Task: approach and landing to SFO RWY 28R (shortened to 7,500 ft for experiment 2)
Initial Condition: trim condition and location with respect to GS and LOC vary each run
Configuration: gear down, flaps full landing, speedbrakes retracted
Weight: maximum landing weight
Ceiling/Visibility: CAVU or 3-mile visibility (depending on run)
Wind: 190/20, 10/20, or 100/10 (depending on run)
Turbulence: moderate **Gusts:** none
Runway: wet, medium braking action, RCC 3/3/3
Procedure:
1) Recover from a possible unstable approach to SFO RWY 28R (full recovery might not be possible)
2) Continue to land on RWY 28R
3) Flare and touchdown meeting, or as close to, desired touchdown criteria as possible
4) Apply thrust reversers and full manual braking
5) Task evaluation ends after the aircraft is fully stopped on the runway

Desired Performance:
1) Longitudinal touchdown: 1,000 –2,000 ft from threshold
2) Lateral touchdown: centerline between main wing gear
3) Sink rate at touchdown: ≤6 ft/s
4) Bring the aircraft to a full stop as quickly as possible

Fig. 5 Flight card for experiments 1 and 2.

All approaches and landings were flown to Runway 28R at San Francisco International Airport for consistency among all the trials. In reality, the runway is 11,870 ft long, 200 ft wide, with a 300-ft displaced threshold. No modifications were made to these dimensions in Experiment 1. However, in Experiment 2, the runway was shortened to 7,500 ft to determine if such a change would affect a pilot's perception of risk. The mean trajectory of one of the experimental conditions is provided in Fig. 4, along with the location of Runways 28L and 28R, and Taxiway C. All trials were flown with moderate turbulence onto a wet runway with medium braking action (runway condition code (RCC) 3/3/3), at the maximum landing weight, and without any automation (including no flight directors). These environmental, runway, and aircraft parameters were selected to increase the difficulty of recovering from an unstable approach and landing the aircraft. The assumption was that go-around criteria developed under these extreme conditions would be conservative and applicable to more favorable conditions in reality.

As the aircraft was always below 1,000 ft for all experiments, it was always in its landing configuration (gear down and flaps at the correct setting for the given aircraft type). Pilots were able to use localizer and glideslope error indicators on the PFD as well as the precision approach path indicator (PAPI) at RWY 28R at their own discretion. In addition, in order to time the flare of the aircraft, standard radio-altitude call-outs began at a main landing gear height of 50 ft and repeated in decrements of 10 ft until touchdown. Pilots were instructed to use maximum manual braking in order to bring the aircraft to a full stop on the runway as quickly as possible.

Touchdown performance metrics were developed using subject matter expert (SME) input and published recommendations and guidance such as the 2017 Flight Safety Foundation Go-Around Decision Making Report [1] and

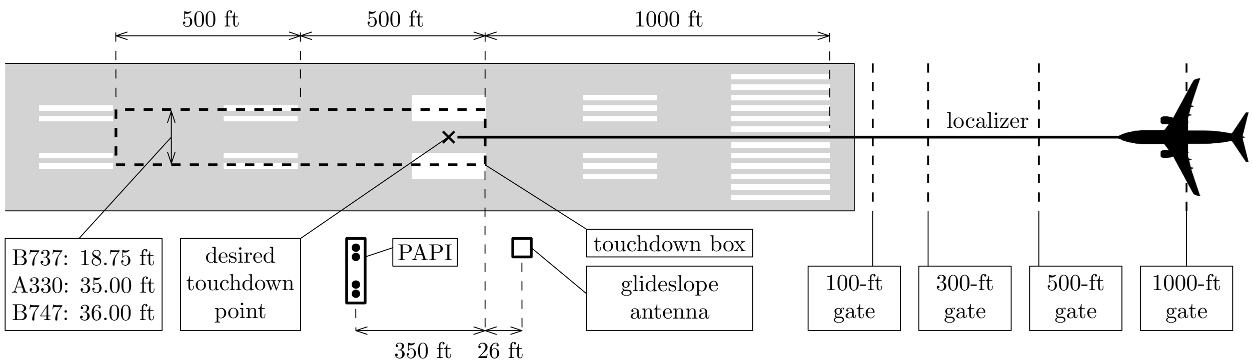

Fig. 6 Touchdown zone definition.

FAA Advisory Circular 91-79A [4]. The selected metrics were representative of what would be considered a routine landing in normal operations. One metric of interest was the touchdown point. Landing too close to the runway threshold increases the risk of an undershoot, and landing too far down the runway increases the risk of a runway overrun. Additionally, large deviations from the runway centerline increase the risk of a runway veer-off. Based on these considerations, a touchdown box was defined to bound the acceptable area for the aircraft to touch down. The touchdown box began at 1,000 ft past the threshold and was 1,000 ft long. The width of the box was equal to the width of the main landing gear of the aircraft: 36, 35, or 19 ft for the B747-400, A330-200, or B737-800, respectively. The touchdown box is illustrated in Fig. 6. The other performance metric of interest was the rate of descent, or sink rate, at touchdown. An unstable approach with high energy could potentially lead to a hard landing. A sink rate threshold of 6 ft/s was used for the touchdown criteria.

Crews comprised of a captain and a first officer from the same airline flew all approaches and landings. Both pilots alternated as the pilot flying (PF) and pilot monitoring (PM) in between sessions. This allowed for objective data and subjective evaluations from both perspectives. The flight card for the approach and landing task is provided in Fig. 5.

2. Independent Variables

Experiment 1 used the following independent variables: gate height (100, 300, or 500 ft), glideslope deviation (0, 0.5, 0.75, or 1.5 dots), localizer deviation (0, 0.5, 0.75, or 1.5 dots), rate of descent (1,000, 1,250, or 1,500 ft/min), and target approach speed deviation (+0, +10, and +20 KIAS). The approach parameters had different levels at each gate height. Experiment-2 independent variables were selected to verify the findings of Experiment 1 [5]. More specifically, the independent variables were selected to further explore the effect of the approach variables on touchdown performance as well as to evaluate the effects of wind and visibility on touchdown performance. Experiment 2 used the following independent variables: gate height (300 or 500 ft), glideslope deviation (0 or 1 dot), localizer deviation (0 or 1 dot), target approach speed deviation (+5, +10, or +15 KIAS), wind (left crosswind at 20 knots, right crosswind at 20 knots,

Table 2 Independent variable settings for experiments 1 and 2.

Gate Height ft	GS Deviation dot	LOC Deviation dot	Rate of Descent ft/min	V_{tgt} Deviation kts	Wind kts	Visibility sm
			Experiment 1			
100	0.0	0.0	1000/1250	+0/+10/+20	tail 10 (fixed)	∞ (fixed)
300	0.0/0.5	0.0/0.5	1000/1500	+0/+10/+20	tail 10 (fixed)	∞ (fixed)
500	0.0/0.75/1.5	0.0/0.75/1.5	1000/1500	+0/+10/+20	tail 10 (fixed)	∞ (fixed)
			Experiment 2			
300	0.0/1.0	0.0/1.0	1200 (fixed)	+5/+10/+15	left 20/right 20/tail 10	∞/3.0
500	0.0/1.0	0.0/1.0	1200 (fixed)	+5/+10/+15	left 20/right 20/tail 10	∞/3.0

or tailwind at 10 knots), and visibility (unlimited or 3 statute miles (sm)). The independent variables for experiments 1 and 2 are summarized in Table 2.

To reduce the size of the final test matrix, other approach parameters, such as the power setting and maximum bank angle, were not included as independent variables. Nonetheless, they were tracked to determine whether they had any significant influence on touchdown performance. In addition, only localizer deviations to the right of the runway were considered, assuming that deviations from the left and right would have similar but opposite effects. This was also more realistic given the parallel runway to the left (RWY 28L) at SFO. In addition, only deviations above the glideslope were considered, as these are more difficult to compensate for compared to deviations below the glideslope. Furthermore, only speed deviations above the target speed (V_{tgt}) were tested; speed deviations below increase the risk of a stall and almost always warrant a go-around.

For both experiments, a test matrix was generated using the custom design-of-experiments (DOE) feature in JMP® [9] to determine the number of runs and scenarios that each pilot would fly during the experiment. A full-factorial design was not used because the allotted time on the simulator would not allow for repeated conditions. The custom DOE option allowed replicates of the corners of the design while maintaining an experiment power of one for all main effects and first order interactions. The test matrix for Experiment 1 had 55 different initial-approach-parameter conditions and a total of 84 conditions with replicates included. The final test matrix had 92 runs, including four training runs and four extra conditions with varied environmental parameters. Data from the extra conditions were used to help develop Experiment 2 (Fig. 1). The test matrix for Experiment 2 had 30 different conditions and a total of 42 runs with replicates included. The final test matrix had 46 runs, including four training runs.

3. Procedures

In Experiment 1, each crew was scheduled for two consecutive days. Experiment 2 was conducted about one year later with a different group of pilots and took one day to complete for each crew. Pilots were provided a briefing document, flight card (shown in Fig. 5), airport diagram, and approach plate before the start of both experiments. On the first day, crews received an extensive pre-briefing, explaining the schedule, task, conditions, and procedures of the experiment. Crews were told that the experiment investigated the effects of different approach parameters on landing performance and were given no specifics with regard to the true nature of the experiments. Crews were informed that they would fly approaches and landings with different initial conditions and that the conditions would be presented randomly. No further details were given regarding the extremity of the conditions. After the briefing, pilots provided their informed consent and filled out a pre-simulation questionnaire. This questionnaire gathered demographic data and information on their airline's current stabilized approach criteria in visual meteorological conditions (VMC), and asked about their satisfaction with those criteria. The experiments started after a simulator safety briefing.

Each crew flew 92 scenarios per day (184 total for Experiment 1 and 92 total for Experiment 2), divided among four one-hour simulator sessions each day. Pilots received breaks outside of the simulator cab between sessions. The length of the breaks was at pilots' own discretion, but was typically 15-30 minutes. Pilots were allowed to take additional breaks at any time during a session if desired. Over the course of the two days of experimentation for Experiment 1, each pilot flew all 92 scenarios in the test matrix. For Experiment 2, each pilot flew all 46 scenarios. Some of the conditions were performed twice. The runs were randomized for every pilot, and pilots rotated between the pilot flying and pilot monitoring roles after each session. The first four runs of the first session were training runs with nominal approach conditions at the different gates, or with different environmental conditions.

Each run started with the aircraft either at the 100-ft gate, 300-ft gate, or the 500-ft gate. The PF manually flew each approach and landed the aircraft during every run of a session. The initial approach and environmental parameters were called out to the pilots by the experimenter before the start of each run to make sure that the pilots were completely aware of the situation. The initial conditions called out by the experimenter also corresponded to the aircraft state variables that would be checked in real life to determine approach stability. After pilots confirmed that they were ready, the experimenter counted down to the initiation of the run. In Experiment 2, both pilots were also instructed to independently press the autopilot disconnect button to indicate the moment at which they normally would have performed a go-around, but were not required to do so for a given run, indicting that they would not have performed a go-around. The PF was asked to meet the touchdown criteria as closely as possible and then to use maximum manual braking and full reverse thrust to bring the aircraft to a complete stop on the runway as quickly as possible. The pilot monitoring was allowed to provide call-outs to assist the PF, as per their airline policy or personal preference. After the aircraft had come to a complete stop, the simulation was re-positioned for the next run, and the pilots answered their post-run questionnaires on tablet computers (see Section III.B.4). Pilots were instructed not to discuss their post-run questionnaire responses with one another.

After completing all simulator runs, pilots filled out a post-simulation questionnaire. This questionnaire asked about the pilots' preferred stabilized approach criteria based on their experiences during the experiment and about which factors most influenced their decision to go around. Finally, each crew received a debriefing providing more details with regard to the true nature of the study.

4. Dependent Measures

Three main objective dependent measures specifying landing performance were recorded and analyzed: longitudinal and lateral touchdown location (x_{td} and y_{td}) as well as the sink rate at touchdown ($\dot{h}_{td}$). These measures related directly to the landing performance criteria the pilots were required to meet as shown in Fig. 5. The touchdown point was defined as where the center of the main landing gear touched the runway with respect to the longitudinal location of the glideslope antenna and the centerline of the runway (Fig. 6). When multiple touchdowns were recorded, the maximum longitudinal distance and the maximum sink rate out of all touchdowns were used. The lateral distance always corresponded with the maximum longitudinal distance. For most cases, this meant that for multiple touchdowns, the sink rate used belonged to the first touchdown, and the longitudinal and lateral touchdown locations used belonged to the last touchdown. The go-around altitude, h_{ga}, was recorded at the time of the autopilot disconnect button press.

In addition to the objective measures, seven subjective dependent measures were recorded using a questionnaire administered on a tablet computer at the conclusion of each run. Pilots first rated their perceived workload, fatigue, and risk of the previous landing (in that order) on a 20-point scale, by moving a slider bar with their finger. Only the ends and midpoints of the slider bars were marked with "low," "average," and "high." Next, pilots were asked if they had pressed the autopilot disconnect button during the run indicating that a go-around should have been performed in their opinion at that time. If they responded with a yes, the pilots indicated which factors influenced their go-around decision by selecting from a list that included the following options: slow, fast, low descent rate, high descent rate, below glideslope, above glideslope, localizer deviation, power setting, bank angle, wind, visibility, turbulence, runway length, and runway condition. It was also possible to select multiple factors for each run. Finally, the last two questions asked the pilot if his/her decision to go around would have been different if the runway was longer or if the braking action was better. Note: both the PF and the PM filled out the post-run questionnaire, resulting in two sets of subjective data for each run.

Additional dependent measures were collected, but only those used in this paper are presented here.

5. Key Findings

Results of Experiment 1 are provided in [5]. Key findings from this experiment included the following.

1) Equivalent approach states at 300-ft and 500-ft gates have similar effects on touchdown performance.
2) An unstable approach state at 100 ft AGL significantly degraded touchdown performance compared to unstable approach states at 300 ft or 500 ft; thus a 100-ft go-around gate was deemed too low.
3) V_{tgt} and localizer deviations at the starting approach gate had the strongest influence on perceived risk.
4) A target speed deviation of +20 KIAS often results in idle thrust usage during the approach.

The key findings listed above resulted in proposed go-around criteria provided in Table 3. Because no statistically significant difference in touchdown performance was found between a starting gate of 300 ft and 500 ft, the FSF recommendation for a go-around gate of 300 ft was adopted in the proposed criteria. In this study, a distinction was made between a go-around gate and a stabilized approach gate. The go-around gate is a final, absolute check, such that if the aircraft is outside of the criteria at this gate, a go-around shall be performed. In accordance with the FSF recommendations, the criteria should also be checked at the 1,000-ft and 500-ft stabilized gates. If the aircraft is outside of the criteria at these gates, the pilot monitoring should verbalize any deviations, and the pilot flying should take the appropriate actions to correct the deviations.

An airspeed deviation of +10 KIAS was selected to be consistent with the FSF recommendations and because the results of Experiment 1 showed little difference in touchdown performance and risk assessments between V_{ref} + 0 and V_{ref} + 10 KIAS. Glideslope and localizer deviations of less than one dot where selected based on established criteria. Lastly, rather than specifying a rate of descent limit, a criterion of no terrain avoidance warning system (TAWS) alert activation was used. This approach was taken for numerous reasons. First, Experiment 1 showed that the rate of descent at the various gate heights had little-to-no effect on touchdown performance. Second, pilots in Experiment 1 were easily able to arrest any initial high rate of descent quickly. Finally, the TAWS is specifically designed to mitigate controlled flight into terrain risk and should alert the pilot of an unsafe descent rate or vertical flight path.

Table 3 Proposed go-around approach criteria.

Criteria at 300 ft		
1	Airspeed	Within 0/+10 of target
2	Glideslope deviation	Less than 1 dot
3	Localizer deviation	Less than 1 dot
4	Rate of Descent	No TAWS activation

The results of Experiment 2 of the study, provided in [6], showed that the environmental factors of wind speed/direction and visibility have a strong effect on touchdown performance. Specifically, wind had a highly significant effect on longitudinal and lateral touchdown point and a significant effect on sink rate at touchdown. In fact, wind had a stronger effect on all three touchdown performance measures than any other variable in the study. Wind and visibility, along with localizer deviation, also had a strong effect on the pilots' perception of workload and risk.

B. Experiment 3 - Pilot Evaluation of Proposed Go-Around Criteria

1. Approach with Landing or Go-Around Task

The primary objective of Experiment 3 was to capture pilot feedback and decision-making with regards to a set of proposed, hypothetical go-around criteria that were developed based on FSF recommendations and the results from Experiments 1 and 2 (Table 3). To this end, pilots flew multiple approaches that were on the borderline of the go-around criteria at 300 ft. Pilots were instructed that they could either execute a go-around or land the airplane on each run, forcing a decision for the borderline cases at 300 ft. The pilots were told to execute a go-around if either the aircraft was outside of the go-around criteria at 300 ft, or if either pilot was uncomfortable with the approach; otherwise, they could decide to land.

A secondary objective of this experiment was to assess the crew's awareness of the aircraft state on the approach. To combine the primary and secondary objectives, each approach began stable at 1,000 ft; then, when the aircraft was between 500 and 300 ft on the approach, it was forced unstable. The idea was for the aircraft to be unstable below 500 ft, but to give the pilots a reasonable chance of re-establishing a stable approach by 300 ft. This method has a few objectives: 1) assess the ability of the pilots to detect the instabilities, 2) evaluate the pilots' acceptance of making corrections below 500 ft down to the 300-ft go-around gate, and 3) evaluate the pilots' acceptance of executing the go-around at the 300-ft gate.

For each run, the aircraft started in exactly the same position: at the 1,000-ft gate and on the glideslope and localizer, with an airspeed of V_{tgt} and a sink rate of 900 ft/min. Visibility was always unlimited and the aircraft always experienced a 10-knot tailwind. Furthermore, all approaches were flown to the artificially shortened version of RWY 28R at San Francisco International Airport, which was 7,500 ft long, 200 ft wide, with a displaced threshold.

The instabilities were generated below 500 ft using custom simulator code. Three types of instabilities were introduced: high airspeed, a localizer deviation to the right, and a glideslope deviation forcing the aircraft to become too high. A rate of descent deviation was not included because the high-on-glideslope condition would force a high rate of descent to recapture the glideslope. The initiation and severity of the instabilities were selected such that the aircraft would become unstable below 500 ft, but could be feasibly re-stabilized by the 300-ft gate. Initiation heights and methods for generating the instabilities are summarized in Table 4, and were selected using SME input and fine-tuned through hundreds of trials by a type-rated pilot and the experiment designers. An example aircraft trajectory with a lateral instability and a landing is shown in Fig. 7.

During each run, the crew had to decide whether to go-around or land. If the crew chose to perform a go-around, then the pilots were instructed to follow their company's go-around standard operating procedures and climb towards 3,000 ft. The simulation would be terminated at 2,000 ft so that the scenario was not unnecessarily prolonged. If the crew chose to land, they were instructed to land the aircraft meeting the prescribed touchdown performance criteria as in experiments 1 and 2 (Fig. 6), then bring the aircraft to a full stop on the runway as quickly as possible by using reverse thrust and maximum manual braking. Fig. 8 provides the flight card for the approach and landing or go-around.

Table 4 Induced instabilities.

Parameter	Target Deviation	Initiation Height (ft AGL)	Method
Airspeed deviation	V_{ref} + 20	450	wind shift to 45 KTS headwind, 5-second duration
Localizer deviation	1 dot	500	Boeing: 36 knot crosswind for 5 seconds; Airbus: lateral shift 1 dot
Glideslope deviation	1 dot	360	altitude freeze for 2 seconds

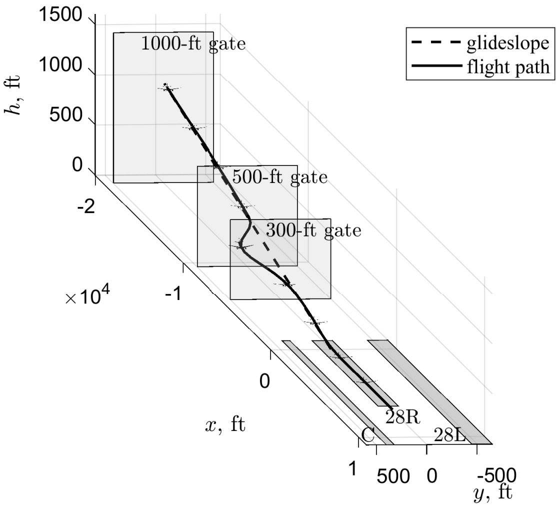

Fig. 7 Mean B737 trajectory with lateral instability for Experiment 3.

Aircraft: B737-800 and A330-200
Task: approach and landing with possible go-around to SFO RWY 28R
Initial Condition: trimmed and stable on GS and LOC at 1,000 ft
Configuration: gear down, flaps full landing, speed brakes retracted
Weight: maximum landing weight
Ceiling/visibility: ceiling and visibility unlimited
Wind: 100/10 **Turbulence:** moderate **Gusts:** none
Runway: short and wet, medium braking action, RCC 3/3/3
Traffic: no traffic or departing traffic on RWY 28L (depending on run)
Procedure:

1 Recover from a possible unstable approach to SFO RWY 28R (full recovery might not be possible)
2 Apply go-around criteria and (a) continue to land on RWY 28R or (b) go-around
3a Flare and touchdown meeting, or as close to, desired touchdown criteria as possible
3b Execute the go-around
4a Task evaluation ends after the aircraft is fully stopped on the runway
4b Task evaluation ends when reaching the missed-approach altitude of 2,000 ft

Desired performance: When continuing to land:

1) Longitudinal touchdown: 1,000 –2,000 ft from threshold
2) Lateral touchdown: centerline between main wing gear
3) Sink rate at touchdown: ≤6 ft/s
4) Bring the aircraft to a full stop as quickly as possible

Fig. 8 Flight card for Experiment 3.

2. Independent Variables

The primary independent variables were the three induced instabilities provided in Table 4. Each of the primary independent variables had two possible settings: stable or unstable. Stable meant that an instability was not forced in the scenario, whereas unstable meant that the simulator code to produce an instability was executed in accordance with the methods described in Table 4. In addition, a traffic condition was included as a fourth independent variable. The traffic condition was a binary variable that determined whether or not a Boeing 747 aircraft was taking off from the parallel runway during the approach. This variable was included to determine whether or not traffic in close proximity of the landing aircraft factored into a pilot's go-around decision-making process. Using the four independent variables, a test matrix was generated using a full-factorial design with two repetitions per independent variable combination. This resulted in 16 experimental conditions and 32 data-collection runs per pilot.

3. Procedures

The same crew that participated in Experiment 2, participated in Experiment 3 the following day. For Experiment 3, pilots were provided a briefing document, a flight card (Fig. 8), and a laminated card with the proposed go-around criteria. The pilots were briefed on the schedule, procedure, and go-around criteria. It was stressed to the pilots that the pilot monitoring should check the criteria at 500 ft, and either pilot should call for a go-around at 300 ft if the aircraft was outside of the criteria or if he/she felt uncomfortable with the approach.

Following the briefing, each crew flew 72 approaches divided among four 1-hour simulator sessions. Each pilot flew four training scenarios followed by the full test matrix. The runs were randomized for every pilot, and the pilots alternated between pilot-flying and pilot-monitoring roles between each session. The first four runs of the first session were training runs with nominal approach conditions. The pilots were asked to land the first two training runs and conduct a go-around at 300 ft for the second two training runs.

The PF flew each approach and the crew either landed or performed a go-around during each scenario. The pilots were told prior to each run whether there was departing traffic on the parallel runway to make sure the pilots were completely aware of the situation. During the approach, the PM was tasked with calling out any deviations from the go-around criteria and then was to call for a go-around if the criteria were not met at the 300-ft go-around gate. If the crew chose to conduct a go-around, they used their company's standard operating procedure and climbed towards the missed approach altitude of 3000 ft. The evaluator terminated the simulation as the aircraft passed through 2000 ft. If the crew chose to land, the PF was asked to meet the touchdown criteria as closely as possible and then to use maximum manual braking and full reverse thrust to bring the aircraft to a complete stop on the runway. The PM was allowed to provide call-outs to assist the PF, as per their airline policy or personal preference. After the aircraft had come to a complete stop, the simulator was repositioned for the next run, and the pilots completed their post-run questionnaires on their tablets (see Section III.B.4).

After completing all simulator runs, pilots filled out a post-simulation questionnaire. This questionnaire asked about the pilots' preferred stabilized approach criteria based on their experiences during the experiment, and about which factors influenced their decision to go around the most. Finally, each crew received a debriefing providing more details about the true nature of the experiment.

4. Dependent Measures

Different dependent measures were captured depending on whether the crew performed a go-around or landed during each run. If the pilots chose to go around, the go-around altitude h_{ga} was determined. The go-around initiation point was defined as the point 2 seconds before the point with the maximum throttle increase over 1 second. If the run ended in a landing, three main objective dependent measures specifying the landing performance were recorded and analyzed: longitudinal and lateral touchdown location (x_{td} and y_{td}) and sink rate at touchdown ($\dot{h}_{td}$). These measures related directly to the landing performance criteria pilots had to meet (see Fig. 8). Data capture for these variables occurred when the main gear touched the runway. When multiple touchdowns were recorded, the same procedure as in experiments 1 and 2 was used to determine the overall touchdown performance parameters.

Subjective dependent measures were recorded using a questionnaire administered on a tablet computer at the end of each run. Pilots first rated their perceived workload, fatigue, and risk of the completed landing or go-around (in that order) on a 20-point scale by moving a slider bar with their fingers. Only the ends and midpoints of the slider bars were marked with "low," "average," and "high." Next, pilots were asked about the acceptability of the 300-ft decision height and the go-around criteria from an operational safety point of view. Once again, a slider bar with a 20-point scale was used. The low end of the scale was marked as "clearly unacceptable," the middle as "indifferent," and the high end as "clearly acceptable." Next, the pilots were asked if they performed a go-around. If they responded with a yes, three questions followed. The first asked about the factors influencing their decision. The factors were selected from a list including the following: slow, fast, low descent rate, high descent rate, below glideslope, above glideslope, localizer deviation, power setting, bank angle, wind, visibility, turbulence, runway length, and runway condition. Then, the second and third questions asked the pilot if his/her decision would have been different if the runway was longer or the braking action was better. If the pilot answered no to the question about whether a go-around was conducted, he/she was asked if the go-around criteria were met at 300 ft. The purpose of this question was to gauge whether the pilot was aware of the aircraft approach state at the 300-ft gate. Note: both the pilot flying and the pilot monitoring filled out the post-run questionnaire, resulting in two sets of subjective data for each run.

Again, additional dependent measures were collected, but only those used in this paper are presented here.

5. Key Findings

Results of Experiment 3 are provided in [7]. The proposed criteria performed well, and most pilots would find the criteria acceptable with some minor adjustments. Key findings were:

1) The objective data suggest that a 300-ft gate is viable with some criteria adjustments such as adding an engine spooled parameter or rate of descent threshold. In general, pilots rated the acceptability of the criteria and the proposed 300-ft go-around gate high at the end of each run. However, 40% of the pilots were uncomfortable with the 300-ft gate after completing the experiment. Placing more emphasis on checking approach stability and using active call-outs at 1000 ft and 500 ft above ground level might make more pilots comfortable using a 300-ft go-around gate. Additional training might be needed to reinforce the concept of having two stabilized approach gates and a go-around gate.
2) The most important factors that drove go-around decision making during the experiment were airspeed and localizer deviation.
3) Allowing for momentary deviations from a stabilized approach should be considered.
4) The acceptability of the criteria is highly dependent on each pilot's risk tolerance.

IV. Combined Results

Results from the three individual experiments can be found in [5], [6], and [7]. This section provides results determined from the combined data of the different experiments. A summary of the data collected in each experiment of the study is shown in Tables 5 and 6. Table 5 summarizes the data collected in the simulator output files. During the analysis, data describing the state of the aircraft at the 500-ft, 300-ft, and 100-ft gates was particularly important for the analysis performed. As indicated in the table, some of the data was collected as the starting state of the run. Otherwise, for some runs the data was collected as a snapshot, meaning data were interpolated at the 300 or 100 ft gates. Table 6 summarizes the questionnaire data collected using a tablet after each run.

Table 5 Summary of performance data collected in the experiments.

	Experiment 1	Experiment 2	Experiment 3
Time series data for entire run	All runs	All runs	All runs
Touchdown performance	All runs	All runs	Only if landing was performed
Aircraft state at 100-ft gate	• Starting state for 100-ft runs • Snapshot for runs starting at the 300-ft and 500-ft gate	Snapshot for all runs	Snapshot if a go around was not performed before reaching this gate
Aircraft state at 300-ft gate	• Starting state for 300-ft runs • Snapshot for runs starting at the 500-ft gate	• Starting state for 300-ft runs • Snapshot for runs starting at the 500-ft gate	Snapshot if a go around was not performed before reaching this gate
Aircraft state at 500-ft gate	Starting state for 500-ft runs	Starting state for 500-ft runs	Snapshot if a go around was not performed before reaching this gate
Go-around height	Not available	Height of A/P disconnect button press	Calculated using time series data

Table 6 Summary of questionnaire data collected in the experiments.

	Experiment 1	Experiment 2	Experiment 3
Workload, fatigue, and risk scores	All runs	All runs	All runs
Input on whether go around was conducted	All runs	All runs	All runs
Go-around height	Available for runs that the pilot indicated that a go around should have been performed	Not available - captured from performance data	Not available - captured from performance data
Reasons for conducting go-around: aircraft state	If go around was selected	If go around was selected	If go around was selected
Reasons for conducting go-around: environmental parameters	Not available	If go around was selected	If go around was selected
Acceptability of proposed gate and criteria	N/A	N/A	All runs
Input on whether the proposed criteria were met at the 300 ft gate	N/A	N/A	Only on runs that resulted in a landing
Input on whether the decision to go around would be different if the runway was longer or the condition was better	N/A	If go around was selected	If go around was selected

Data were successfully captured from all participants in the two experiments. Data from the simulator and tablet questionnaires were parsed using MATLAB® and then analyzed using various tools including MATLAB®, R, and JMP® [10, 11]. Statistical analyses were primarily conducted using R and JMP®. Section IV.A presents the results from the landing data and Section IV.B from the go-around data. Finally, Section IV.C presents the results on pilots' acceptability of the current and proposed stabilized approach or go-around criteria.

A. Landing Data

This section uses data from all landings performed in the study. This means all runs of experiments 1 and 2, and runs of Experiment 3 that resulted in a landing.

1. Touchdown Performance

Linear mixed-effects models determined if the aircraft state at the 100-ft, 300-ft, or 500-ft gates (independent variables) introduced significant differences in touchdown performance. Separate models were fit for each gate and performance measure (x_{td}, y_{td}, and $\dot{h}_{td}$). The models included localizer deviation (ΔLOC), glideslope deviation (ΔGS), rate of descent (ROD), target approach speed deviation (ΔV_{tgt}), wind, visibility, and aircraft type as fixed effects. In addition, the interactions of ΔLOC, ΔGS, ROD, ΔV_{tgt}, wind and visibility with aircraft type were included as fixed effects. Pilot was used as the random effect. The models included random intercepts only; that is, no random slopes were introduced. The random effect of pilot explained around 20%, 7%, and 13% of the total variance for the models of the longitudinal and lateral touchdown points, and sink rate at touchdown, respectively.

The aircraft state at the start of the runs was used in experiments 1 and 2. In addition, snapshots of the aircraft state at gates below the starting gate were used for all experiments (see Table 5). Few outliers were present in the data, as assessed by boxplots, so they were kept in the analysis. Assumptions of linearity, homoskedasticity, and normality of residuals were checked visually using scatter plots and Q-Q plots of the residuals for each model. For the absolute lateral touchdown point, $|y_{td}|$, and the sink rate at touchdown, $\dot{h}_{td}$ the model assumptions of homoskedasticity and normality of residuals were not met due to the fact that these measures were highly positively skewed. These dependent measures were transformed to make them more normally distributed using a Box-Cox transformation [12]. All the models met the assumption of homoskedasticity and normality of residuals after the transformations were applied. No other violations of the assumptions were detected.

The mixed-effects models were progressively built up by adding the different fixed effects and interactions one-by-one, starting with the intercept-only model. Likelihood ratio tests between the models with and without a fixed effect or interaction determined the significance of that effect. Effects that were not significant were removed from the model. Table 10 in the appendix provides the results of the likelihood ratio tests. The fixed factors were added to the model in the order from top to bottom in Table 10.

Figs 9 and 10 plot the touchdown performance parameters as a function of aircraft state at the 100-ft, 300-ft, and 500-ft gate. Each dot represents a single run; however, not all data points are plotted for readability, but a random selection. The different colors represent the different aircraft types (B737, B747, or A330). The aircraft type is not

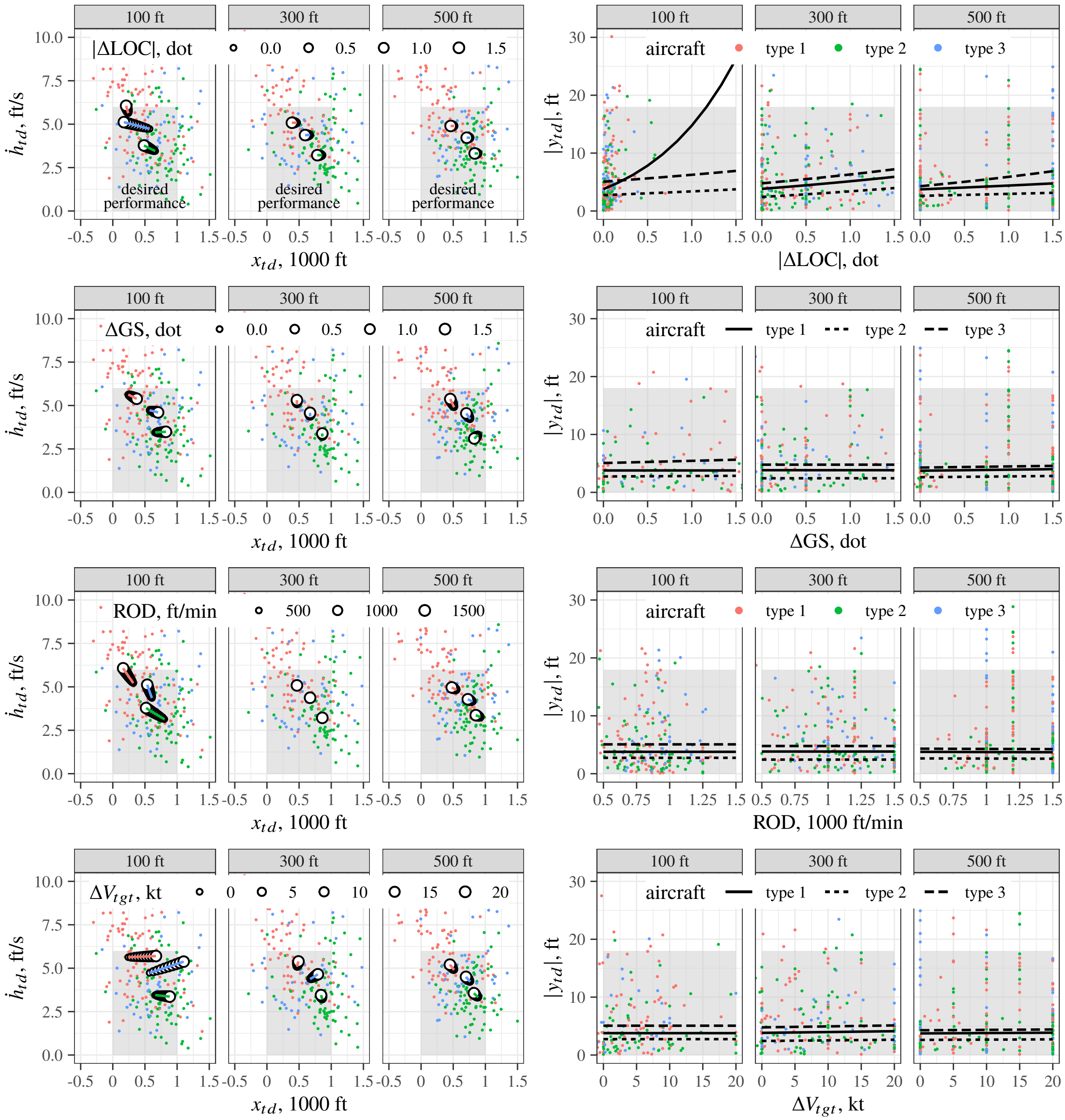

Fig. 9 Effects of the aircraft state independent variables on touchdown performance.

identified specifically so that performance differences between aircraft make and models are not implied. The plots on the left in Figs 9 and 10 depict longitudinal touchdown point versus sink rate at touchdown. In these plots, the mixed-effects model predictions as function of a particular independent variable are presented by white dots. The dots increase in size over the range of the dependent variable. The plots on the right in Figs 9 and 10 depict the lateral touchdown point as a function of a particular independent variable. The model predictions in these plots are provided by the different black lines. To calculate the model predictions, the independent variable presented in the plot was varied over the range depicted and the other independent variables were kept at nominal values (ΔLOC = 0, ΔGS = 0, ROD = 1000, ΔV_{tgt} = 0, wind = 10 tail, and visibility = unlimited). Gray areas indicate the regions of adequate performance as defined in Fig. 5.

The top row of plots in Fig. 9 depicts touchdown performance as a function of localizer deviation. Localizer deviations at the 500-ft, 300-ft, and 100-ft gates introduced significant differences in the longitudinal and lateral touchdown point, and, at 100-ft, also in the sink rate at touchdown (Table 10). Larger localizer deviations resulted in pilots touching down significantly earlier and resulted in significantly larger lateral touchdown deviations. In addition, localizer deviations at the 100-ft gate resulted in significantly higher sink rates at touchdown. The interaction of localizer deviation at 500 ft with aircraft type was also significant for the lateral touchdown point and the interaction of localizer deviation at 100 ft with aircraft was significant for the longitudinal and lateral touchdown points. For aircraft type 1, the effects from localizer deviation at 100 ft on the longitudinal touchdown point were less significant and on the lateral touchdown point more significant, compared to the other aircraft types.

Touchdown performance as a function of glideslope deviation at the different gates is shown in the second row of plots in Fig. 9. Glideslope deviations at the 500-ft gate introduced significant differences in the longitudinal and lateral touchdown points and the sink rate at touchdown. In addition, the interaction with aircraft type introduced significant differences in the sink rate at touchdown. Glideslope deviations at 300 ft introduced significant differences in the sink rate at touchdown only. The interactions of glideslope deviation at the 100-ft gate with aircraft were significant for all touchdown performance parameters, but only the main effects on longitudinal touchdown point and sink rate at touchdown were significant. Most notably, when the aircraft was above glideslope at the 100-ft gate, pilots touched down significantly later. Furthermore, being above glideslope at 300 and 500 ft resulted in higher sink rates at touchdown.

The third row of plots in Fig. 9 depicts touchdown performance as a function of rate of descent. Rate of descent at 500 ft introduced significant differences in all touchdown performance parameters. However, rate of descent at 300 ft did not introduce any significant effects. Rate of descent at 100 ft introduced significant differences in longitudinal touchdown point and sink rate at touchdown. The interaction of rate of descent at 100 ft with aircraft type was also significant. Higher rates of descent at 100 ft resulted in pilots landing shorter and with higher sink rates.

Touchdown performance as a function of target speed deviation at the different gates is shown in the bottom row of plots in Fig. 9. V_{tgt} deviation at the 500- and 300-ft gates introduce significant differences in all touchdown performance parameters (see Table 10). The interaction of V_{tgt} deviation at the 300-ft gate with aircraft had a significant effect on the longitudinal touchdown point. V_{tgt} deviation at 100 ft and its interaction with aircraft type introduced significant differences in the longitudinal touchdown point. The interaction of V_{tgt} deviation at 100 ft with aircraft type introduced significant differences in sink rate at touchdown. Higher V_{tgt} deviations at 100 ft resulted in pilots touching down significantly longer and, for aircraft type 3, with significantly higher sink rates as well.

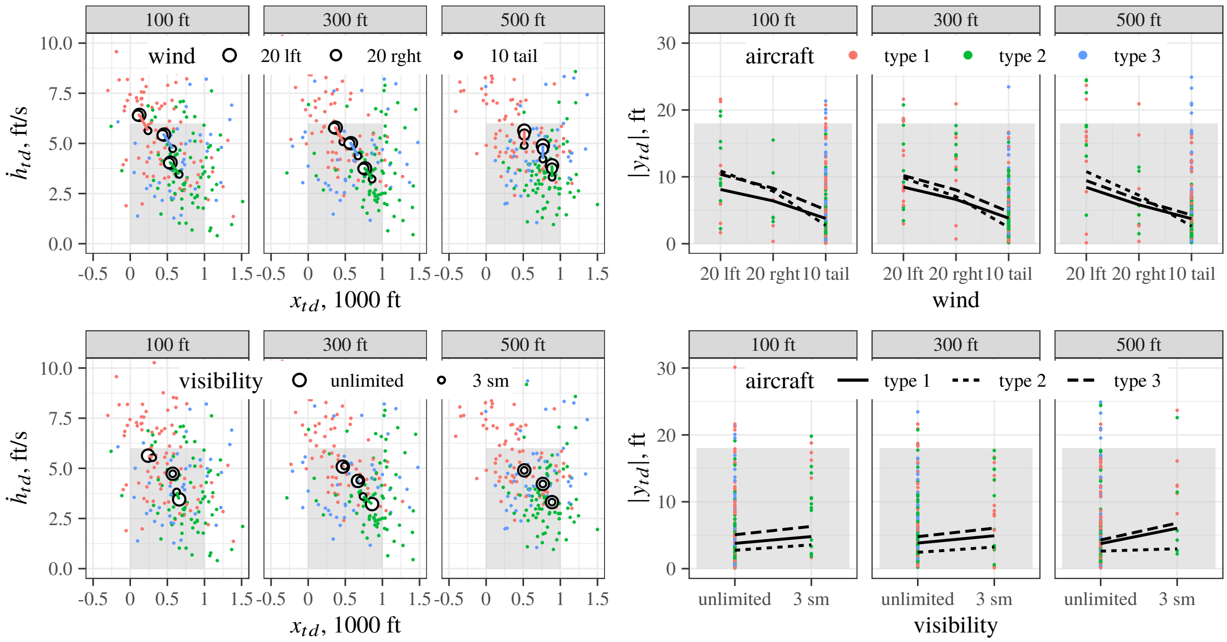

Fig. 10 Effects of the environmental independent variables on touchdown performance.

The effect of wind on touchdown performance is provided in the top row of plots in Fig. 10. Note that the wind conditions prevailed throughout the entire run; however, looking at its effects starting at different gates reveals the pilots' ability to correct for it given the remaining distance from the runway. Wind condition at all gates introduced significant differences in all touchdown performance parameters, with the only exception of the wind condition at the 500-ft gate having no significant effect on the longitudinal touchdown point. In addition, the interaction of wind and aircraft type at all gates introduced significant differences in the lateral touchdown point. With a 20-kt crosswind, pilots touched down harder and sooner, and with larger lateral deviations from the centerline.

Touchdown performance as a function of visibility is provided in the bottom row of plots in Fig. 10. Again, note that the visibility conditions prevailed throughout the entire run. The visibility condition at the 300-ft gate introduced significant differences in all touchdown performance parameters. The visibility condition at the 500- and 100-ft gates introduced significant differences in the lateral touchdown point only. The interaction of visibility at the 500-ft gate with aircraft type introduced significant differences in lateral touchdown point, while the interaction of visibility at the 300- and 100-ft gates with aircraft introduced significant differences in the longitudinal touchdown point and sink rate at touchdown. The reduced visibility at the 300- and 100-ft gates resulted in pilots landing slightly longer. The reduced visibility at all gates resulted in larger lateral deviations from the centerline at touchdown.

Note that the aircraft type introduced significant differences in all touchdown performance parameters. This is an expected result given the different sizes and makes of the aircraft types used in this study. Finally, considering effect size determined by the mixed-effects models and as observed in Fig. 9, the significant differences in touchdown performance introduced by localizer deviation, glideslope deviation, rate of descent, and target speed deviation at the 100-ft gate can mostly be considered operationally relevant (with maximum values of around 20 ft lateral touchdown deviation for one aircraft type and 500 ft longitudinal touchdown deviation for the independent variable range in Fig. 9). However, the significant differences introduced by these independent variables at the other gates are most likely not operationally relevant (with maximum values of around 2 ft lateral touchdown deviation and 100 ft longitudinal touchdown deviation). The effects of the environmental variables (Fig. 10) on touchdown performance are not dependent on the gate; that is, the effects are similar irrespective of the starting gate.

2. Perceived Workload, Fatigue, and Landing Risk

After each run, the tablet questionnaire first asked pilots to rate their perceived fatigue, workload, and landing risk for the previous run. Fatigue, workload, and risk were rated on a scale between 0 and 20. A raster plot of risk evaluation for each run as a function of perceived workload and fatigue is shown in Fig. 11. The figure shows the responses of the pilots flying and monitoring combined. Some areas of the plot have no data (most notably a combination of low workload and high fatigue levels), as pilots did not rate any of the runs with certain combinations of fatigue and workload levels. The plot shows that the perception of high landing risk was associated with elevated perceived workload and fatigue. However, workload had a stronger correlation with perceived landing risk than fatigue.

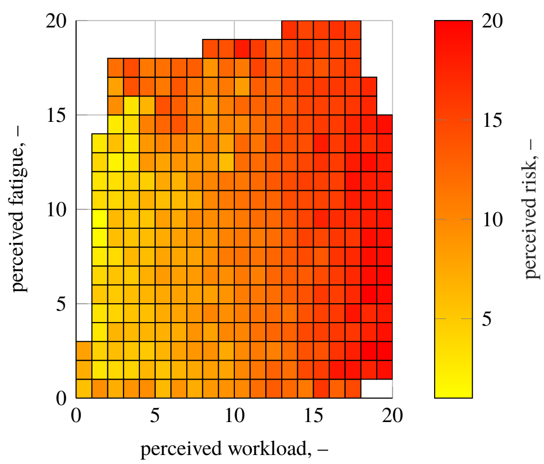

Fig. 11 Perceived risk vs. perceived workload and fatigue (combined PF and PM data).

The responses for perceived workload, fatigue, and risk varied widely between pilots. To demonstrate the variability of the responses by pilots, the scatter plot in Fig. 12 shows the standard deviation and mean of the workload, risk, and fatigue responses for each pilot. The plots show that the responses for workload had means ranging from 1.3 to 14.3 and standard deviations ranging from 0.6 to 5.8. Similarly, the mean responses for risk ranged from 1.3 to 15.0 and had standard deviations ranging from 0.5 to 5.5. The mean responses for fatigue ranged from 0.2 to 15.2 and had standard deviation from 0.1 to 5.4.

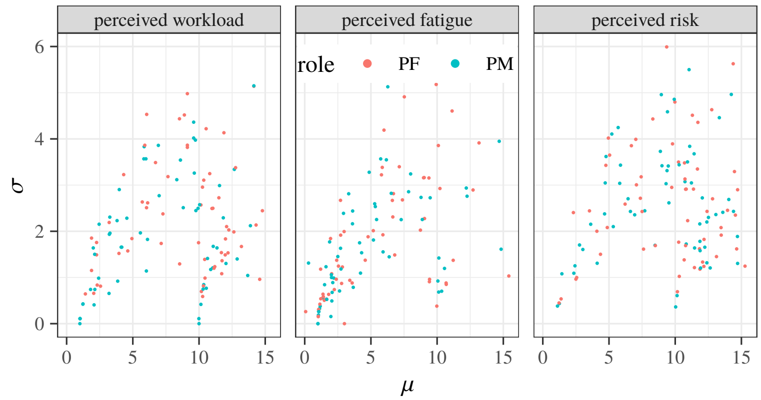

Fig. 12 Standard deviation vs. mean of perceived workload, fatigue, and risk responses by pilot.

Linear mixed-effects models determined if the aircraft state at the 100-ft, 300-ft, or 500-ft gates (independent variables) introduced significant differences in the perceived workload and risk assessment. Separate models were fit for each gate for workload and risk. Again, the models included localizer deviation (ΔLOC), glideslope deviation (ΔGS), rate of descent (ROD), target speed deviation (ΔV_{tgt}), wind, visibility, and aircraft type as fixed effects. In addition, the interactions of ΔLOC, ΔGS, ROD, ΔV_{tgt}, wind and visibility with aircraft type were included as fixed effects. As both the pilot flying and pilot monitoring provided ratings, pilot role and the interaction of pilot role with aircraft were included as fixed effects as well. Pilot was used as the random effect. The models included random intercepts only; that is, no random slopes were introduced. Only the aircraft state at the start of the runs in experiments 1 and 2 was used.

Figs 13 and 14 plot perceived workload and risk, respectively, as a function of aircraft state at the 100-ft, 300-ft, and 500-ft gate. Each dot represents a single run; however, not all data points are plotted for readability, but a random selection. The different colors represent the different aircraft types (B737, B747, or A330). The large variability in responses is apparent from these figures as well. Note that at the 100-ft gate, only rate of descent and target speed deviations were varied (see Table 2).

Fig. 13 depicts the effects of the independent variables on perceived workload. All independent variables at all gates introduced significant differences in workload, except for glideslope deviation and rate of descent at the 300-ft gate (see Table 10). In addition, the interaction of localizer deviation and aircraft at the 500- and 300-ft gates was significant. An increase in localizer and glideslope deviations, rate of descent, and V_{tgt} deviation resulted in higher workload ratings. Furthermore, perceived workload was higher with a 20-kt crosswind compared to a 10-kt tailwind, and higher with a 3-sm visibility compared to unlimited visibility.

The effects of the independent variables on perceived landing risk is provided in Fig. 14. All independent variables at all gates introduced significant differences in pilots' risk assessment, except for glideslope deviation and aircraft type at the 300-ft gate (Table 10). The interaction of localizer deviation and aircraft at the 500- and 300-ft gates was significant, as well as the interactions of glideslope deviation, rate of descent, and visibility with aircraft at the 300-ft

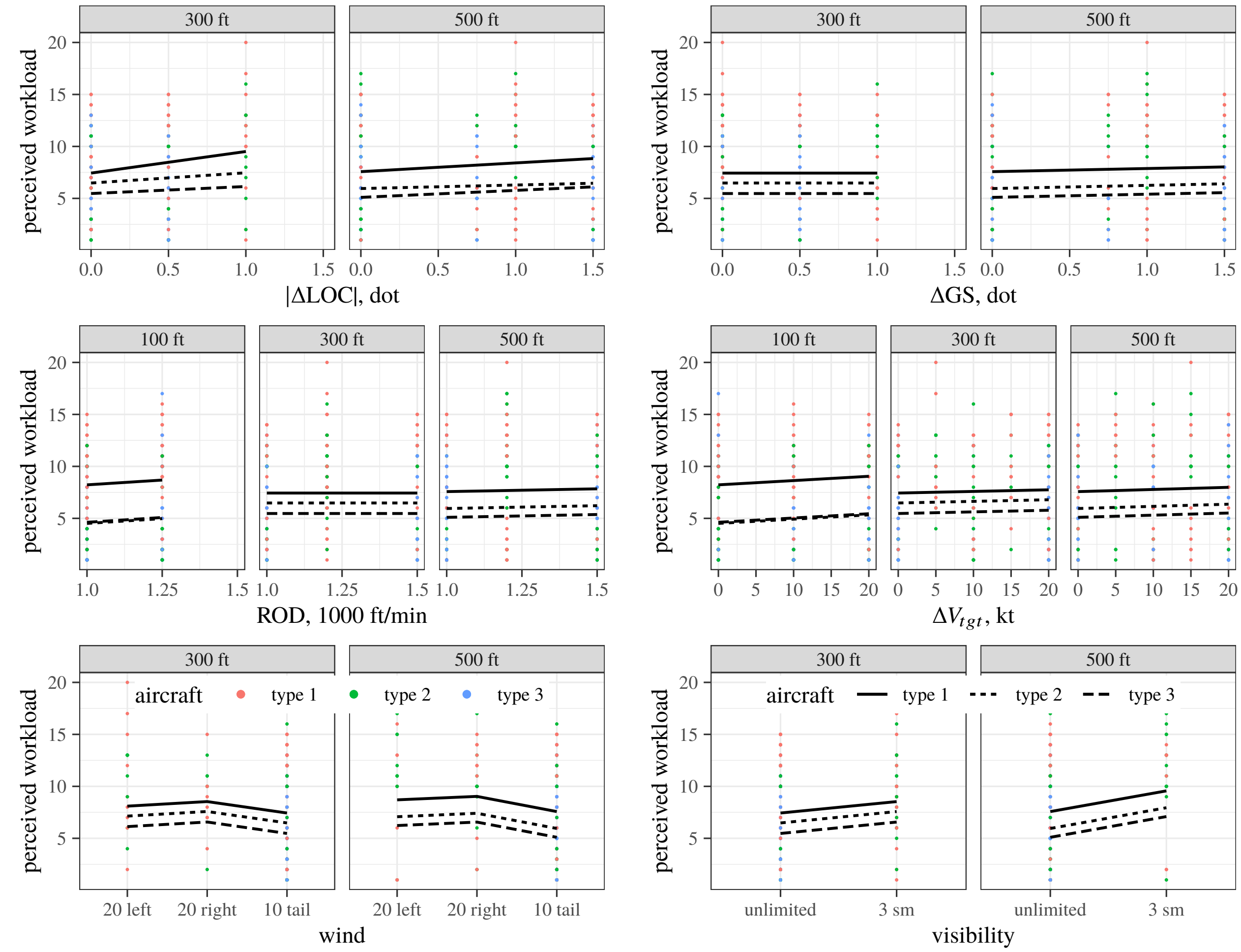

Fig. 13 Effects of the independent variables on workload.

gate. Similar to the workload ratings, an increase in localizer and glideslope deviations, rate of descent, and V_{tgt} deviation resulted in higher risk ratings. Furthermore, pilots' risk rating was higher with a 20-kt crosswind compared to a 10-kt tailwind, and higher with a 3-sm visibility compared to unlimited visibility. Finally, the workload and risk ratings were significantly higher for the pilot flying compared to the pilot monitoring.

B. Go-Around Data

This section uses data from all go-around runs in the study. Only Experiment 3 allowed pilots to go around; however, pilots indicated if they would have gone around and the height of the go-around initiation in the first two experiments.

1. Go-around Initiation Height

The go-around height for each experiment is depicted in Fig. 15. In Experiment 1, the go-around height initiation was the go-around height entered on the tablet questionnaire (see Tables 5 and 6). In Experiment 2, the go-around height was the height AGL when the autopilot disconnect button was pressed. In the first two experiments both the pilot

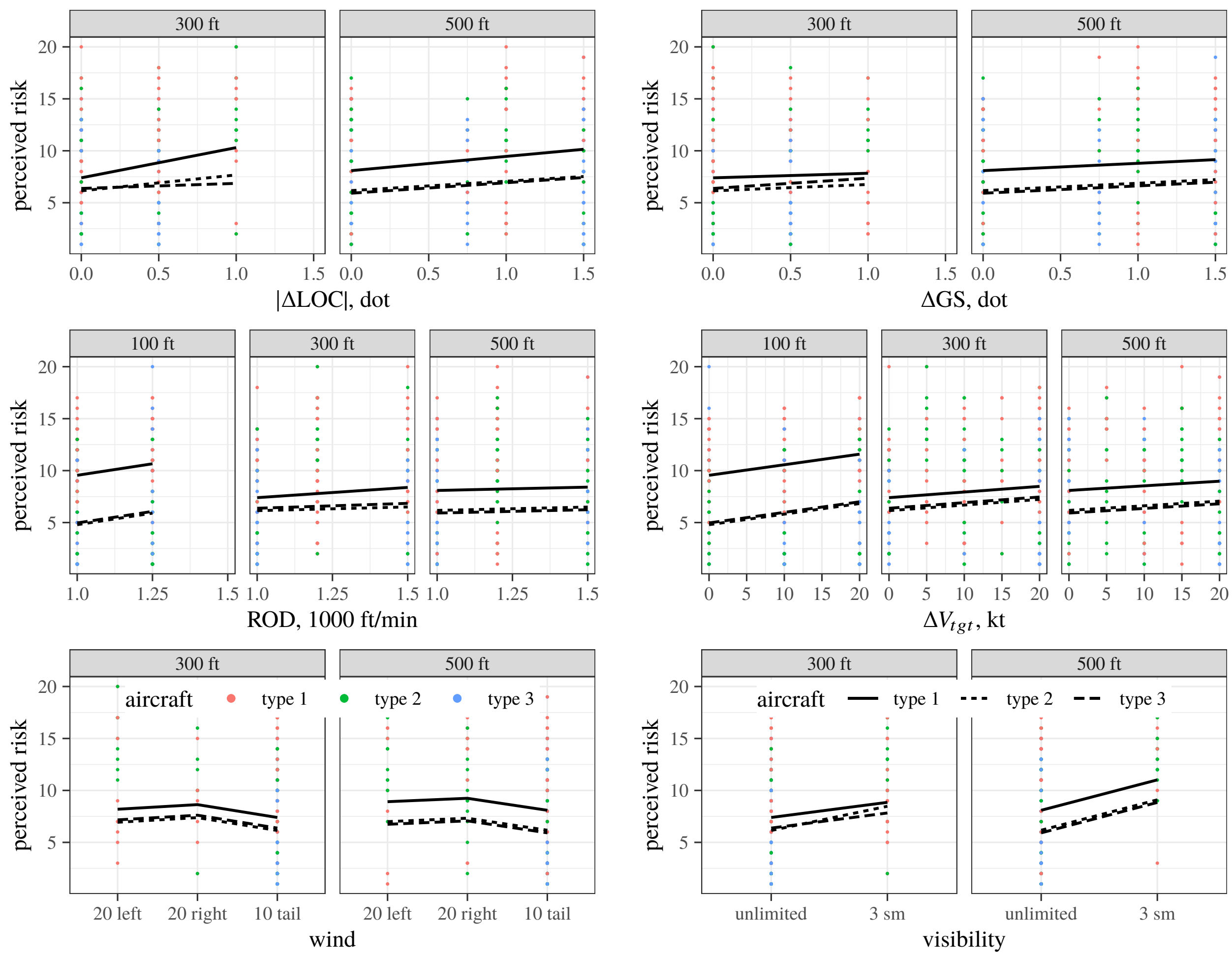

Fig. 14 Effects of the independent variables on landing risk.

flying and the pilot monitoring provided a go-around height in the questionnaire or pressed the autopilot disconnect button. In Experiment 3, the go-around initiation height was the actual height AGL that the go-around was initiated. The PF and the PM indicated a go-around should have been initiated for 985 and 929 out of 3024 runs (33% and 31%), respectively, in Experiment 1. This was 194 and 215 out of 1008 runs (19% and 21%), respectively, in Experiment 2. The PF initiated a go-around in 416 out of 768 runs in Experiment 3 (54%). The go-around rates for all three experiments were significantly higher than the operational go-around rate. This is likely due to a number of factors such as a focus on unstable approaches and go-arounds in the simulation environment and the lack of operational pressures.

Fig. 15 indicates that the distribution of the go-around height was different for each experiment. The go-around initiation height for Experiment 3 typically was the highest while the go-around initiation height for Experiment 2 tended to be the lowest. This was most likely caused by the differences in the scenarios for the different experiments and how the go-around initiation height was recorded. The responses for Experiment 2 probably tended to be lower than for Experiment 1 because for Experiment 1 the pilots provided an answer after the scenario completion rather than giving an input during a run. Therefore, pilots would often state that a go-around would have been initiated at the start of the scenario; however, during Experiment 2, the pilots would have to manually press the autopilot disconnect button to indicate a go-around would be performed which would often be delayed after the start of the scenario. Additionally, in many cases the pilots stated that they forgot to press the button when they would have actually initiated the go-around, so the recorded height of the go-around intended was lower than the pilots intended.

Pilots in the study would often go-around as soon as an instability was encountered. Therefore, the reason that the Experiment 3 go-around initiation heights were higher than the first two experiments was likely caused by differences in the heights that pilots would encounter an instability. In Experiment 1, some scenarios started at 100 ft AGL; meaning that the pilots could not select a go-around height higher than 100 ft. Similarly for Experiment 1 and 2 scenarios that started at 300 ft, the go-around initiation point would be at 300 ft or lower. However, in Experiment 3, all of the instabilities were triggered before 300 ft AGL; therefore, most of the go-arounds were initiated earlier than in experiments 1 and 2.

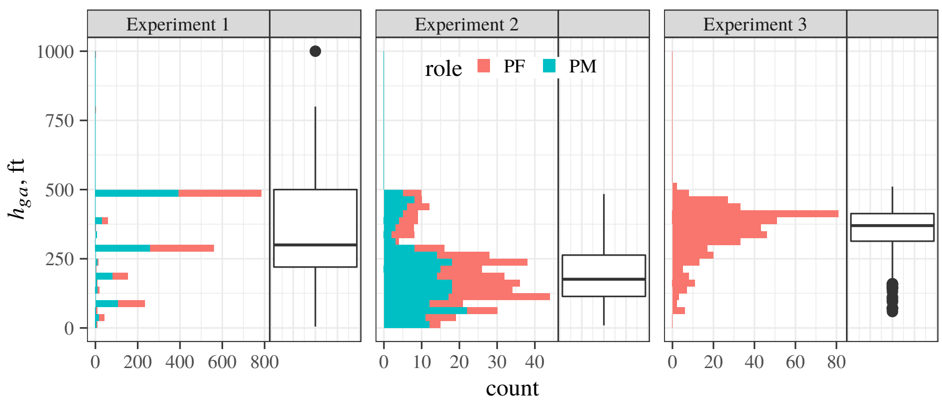

Fig. 15 Distributions of go-around height by experiment.

2. Go-Around Decision

After each run during all experiments, if a pilot responded that a go-around should have been conducted on the tablet questionnaire, he/she was asked to select the reasons for the go around. Table 7 shows the percentage at which each reason for performing a go-around was selected for all experiments. The reasons varied slightly by experiment, probably because of differing starting conditions. LOC deviation and above GS was consistently selected as a reason to go-around. Too fast was selected often in experiments 1 and 3. It was selected less often for Experiment 2 most likely because the starting conditions had less extreme airspeeds. High ROD was consistently selected in the first two experiments, most likely because it was one of the factors in the experiment. The only way a high ROD was generated in Experiment 3 was if the pilot was correcting for being high of the GS. Wind was only a selectable parameter for experiments 2 and 3 and was selected fairly often, highlighting the importance of wind conditions on the go-around decision.

Table 7 Percent of runs reasons for go-around were selected on tablet questionnaire.

	Experiment 1	Experiment 2	Experiment 3
	%	%	%
Too Fast	57.8	24.7	46.2
Localizer Deviation	36.6	50.5	41.1
Above Glideslope	34.8	39.9	31.0
High Rate of Descent	48.4	43.6	24.8
Wind	–	29.9	23.4
Runway Length	–	22.2	14.4
Turbulence	–	5.7	12.5
Runway Condition	–	15.5	12.1
Power Setting	18.5	9.0	7.6
Bank Angle	2.4	14.4	5.5
Below Glideslope	4.8	23.7	4.5
Low Rate of Descent	0.2	3.6	0.6

Experimental factors affecting a pilot's decision to go around were also analyzed through the use of decision trees, also known as Classification and Regression Trees (CART), or partition models. Decision trees are popular in the fields of machine learning, artificial intelligence, and predictive analytics, along with other tree-based models such as random forests. Decision trees are relatively simple, and although other models may provide higher accuracy and predictive power, decision trees were chosen for this analysis in this study for their interpretability over other models such as neural networks, which might have been more difficult to interpret unless a high accuracy was achieved [13, 14].

Decision trees were created using the partition modeling platform in JMP®, which recursively divides the dataset for each experiment based on the relationships between the input and output variables, until the distribution of data points at the terminal nodes, or leaves, is able to predict the output variable. Input variables, or predictors, can be either continuous, such as localizer and glide slope deviations, or categorical, such as wind conditions and visibilities. If a predictor is continuous, then splits are created using a "cutting value," and the dataset is divided into smaller datasets with values below and above the cutting value for the given predictor variable. If a predictor is categorical, then the dataset is divided into two groups of levels, where each group may contain multiple levels of the given predictor. For this analysis, the output variable, or response, was the go-around decision, as indicated by the pilot on the tablet questionnaire following each run, and was treated as a continuous variable so that the mean of the go-around decision at the terminal nodes refers to the percentage of go-arounds conducted in the subset of data determined by the decision tree [14].

Node splitting is performed automatically in JMP® and is chosen such that the difference in the response, the mean of the go-around decision, between the two nodes of the split is maximized. Node splitting is based on the LogWorth statistic, essentially a transformed p-value, and is shown in Eq. (1) below [14].

$$LogWorth = -log_{10}(\text{p-value}) \quad (1)$$

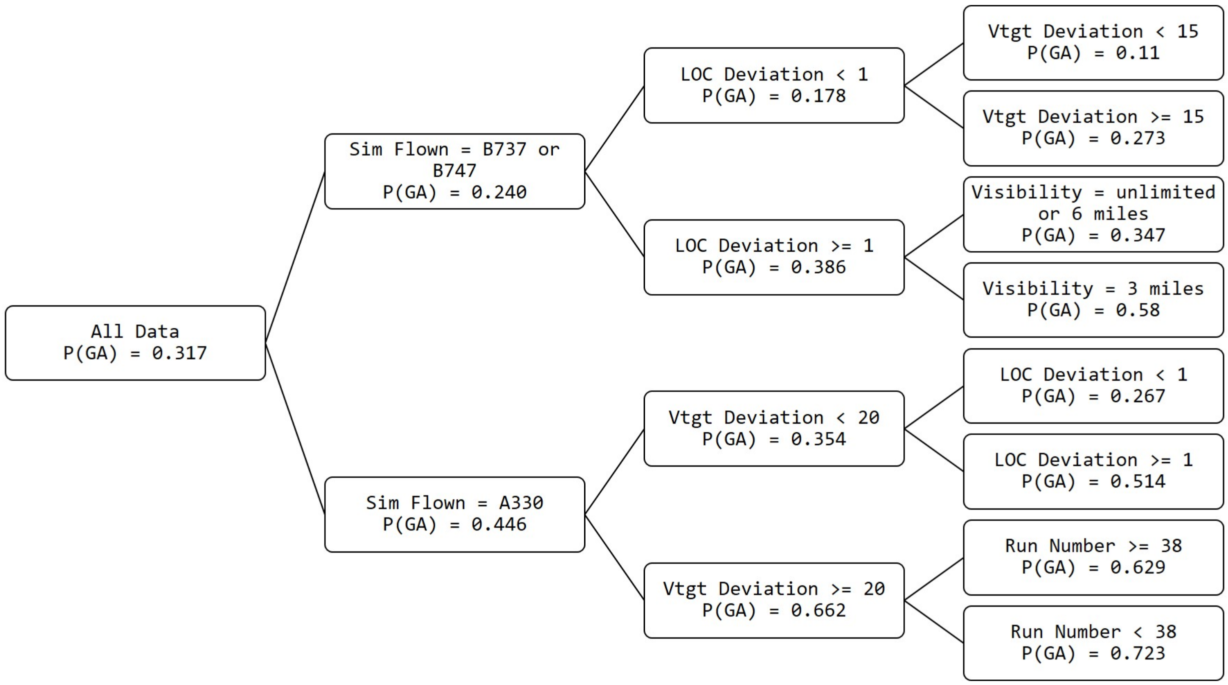

Fig. 16 Decision tree for the go-around decision for the first two experiments.

For this analysis, decision trees were created in JMP® using the experimental factors as the input variables and the go-around decision from the tablet questionnaire as the output variable for the combined dataset from experiments 1 and 2, shown in Fig. 16, and the dataset from Experiment 3, shown in Fig. 17, where deviations were categorical variables instead of numerical. Both datasets used for the decision tree analysis included all runs, regardless of the go-around decision.

The decision tree created by combining the go-around decision datasets from the first two experiments shows that on average, pilots chose to conduct a go-around approximately 31.7% of the time, with a standard deviation of approximately 0.47. However, when the overall dataset was split by aircraft type (as determined automatically by JMP®), the average increased to 44.6% for pilots flying the A330 simulator and decreased to 24.0% for pilots flying the Boeing 737 and 747 simulators. Splitting the datasets further shows that V_{tgt} deviation was more important to A330 pilots and localizer deviation was more important to B737 and B747 pilots in the first two experiments. It can also be seen that other important factors for the pilots include the visibility, where a three-mile visibility increased the probability of a go-around, as well as the run number, where a higher run number reduced the probability of a go-around indicating possible learning and/or fatigue effects.

The decision tree created from the go-around decision dataset from Experiment 3 is different from that of the first two experiments in that deviations are categorical variables as opposed to continuous numerical variables. It can be seen from the Experiment 3 decision tree, shown in Fig. 17, that on average, pilots chose to conduct a go-around approximately 54.0% of the time, significantly higher than in the first two experiments, with a standard deviation

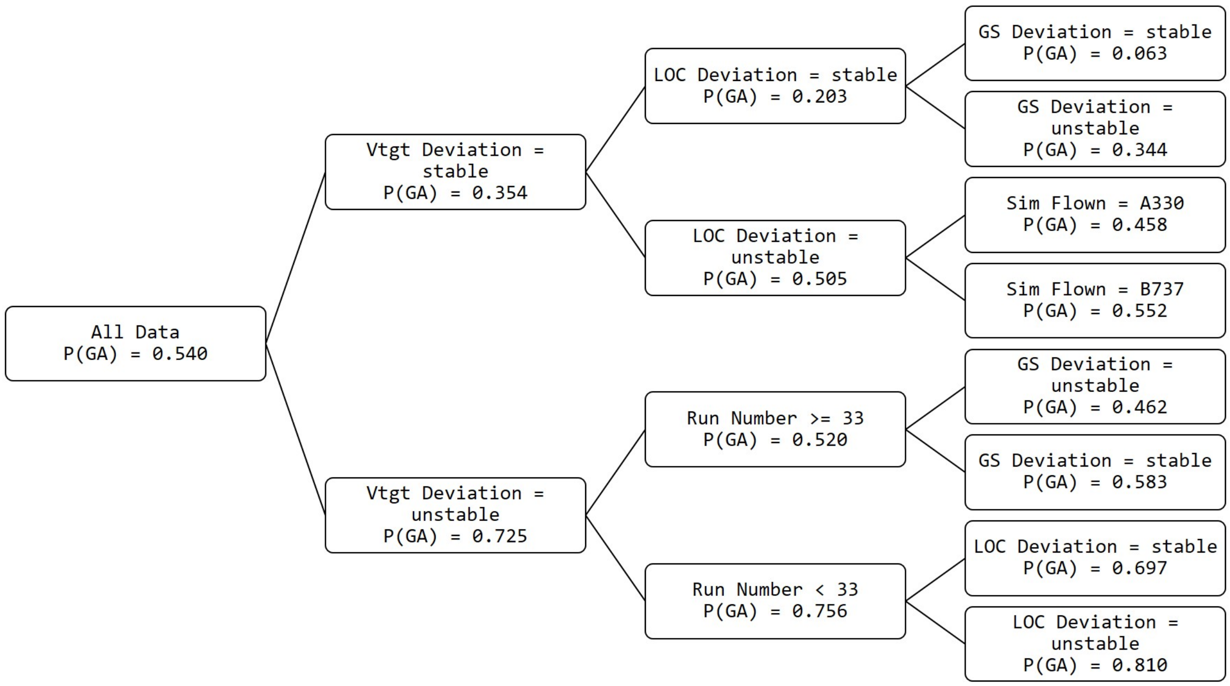

Fig. 17 Experiment 3 decision tree for the go-around decision.

of approximately 0.50. In addition, and in contrast to the first two experiments, the primary deciding factor for the go-around decision was no longer the aircraft type. Instead, it was the V_{tgt} deviation, where an unstable V_{tgt} deviation increased the go-around probability to 72.5% and a stable V_{tgt} deviation decreased the go-around probability to 35.4%. For stable V_{tgt} deviations, the secondary deciding factor was the localizer deviation, and for unstable V_{tgt} deviations, the secondary deciding factor was the run number, where higher run numbers decreased the probability of a go-around, similar to the first two experiments. The simulator flown did have a small effect on the go-around probability for runs where the V_{tgt} deviation was stable and the localizer deviation was unstable, in which case pilots flying the B737 decided to go-around at a slightly higher rate than pilots flying the A330. The B747 simulator was not utilized in Experiment 3.

C. Acceptability of Criteria

In the pre- and post-simulation questionnaires, pilots provided feedback on the acceptability of current stabilized approach criteria (Table 1) and the proposed go-around criteria and go-around gate (Table 3).

In Experiment 1, all but two pilots indicated they were satisfied with their airline's stabilized approach criteria. Post simulation questionnaire responses for what stabilized approach criteria would be acceptable from an operation safety point of view varied widely. The two most common responses for "What parameters have the strongest influence on your decision to go-around rather than continue an approach?" were airspeed and rate of descent.

In Experiment 2, the responses varied significantly when pilots were asked to define their own stabilized approach criteria. When pilots were asked to rank the parameters that have the strongest influence on their go-around decision,

Table 8 Experiment 3 post-sim written questionnaire response summary.

Question	No	Yes	No Preference
Were the proposed criteria acceptable?	10	14	n/a
Were the proposed criteria an improvement over your airline's current criteria?	11	11	2
Would you change the proposed decision height of 300 ft?	12	12	n/a
Would you change the proposed lateral deviation criterion?	20	4	n/a
Would you change the proposed vertical deviation criterion?	17	7	n/a
Would you change the proposed rate of descent criterion?	15	9	n/a
Would you change the proposed airspeed criterion?	17	7	n/a
Would you change the proposed bank angle criterion?	17	7	n/a
Would you change the proposed power setting criterion?	14	10	n/a

configuration was ranked first most often being selected as the top parameter by 12 out the the 24 pilots. Configuration was followed by lateral deviation (selected by five pilots as most important) and vertical deviation (selected by four pilots as most important). Similarly, pilots were asked to rank four environmental parameters (visibility, wind direction and speed, runway condition, and runway length) on their importance when deciding to go around. Nine pilots selected visibility as the most influential, eight selected wind direction and speed, six selected runway condition, and two selected runway length. Eighteen pilots responded that the criteria should be adapted based on environmental parameters. Many of those pilots suggested that the criteria should be more conservative in instrument meteorological conditions (IMC). Through the questionnaire, pilots indicated in their opinion what parameters put a landing at a high risk for a runway excursion. Speed and long landing were the parameters most often mentioned.

After Experiment 3, pilots were asked a series of questions about the acceptability of the proposed criteria and about whether they recommended any changes to them. The responses are summarized in Table 8. The largest number of Yes responses on the parameter to change (other than decision height) was power setting. Pilots also had the most consistent response of how to change the parameter (i.e. add engine spooled requirement/power stable). Other parameters such as allowable dots deviation on the localizer had a variety of responses such as making the criterion 0.5 dots or 1.5 dots. Of the pilots that responded that the proposed criteria were not acceptable and said to add an engine spooled requirement, the responses for other parameters to change was mixed. The other parameter those pilots wanted changed most often was the vertical deviation; however, the value responses were split between 0.5 dot and 1.5 dot. One interesting finding is that the power setting was rarely selected as a trigger for a go-around; however, many of the pilots responded in Experiment 3 that power setting should be part of the criteria.

V. Discussion

The results of the study provided insight into the effect of various approach and environmental parameters on touchdown performance and go-around decision-making. Using this information and the subjective questionnaire data collected, the acceptability of the go-around criteria proposed in Table 3 was evaluated and necessary adjustments to the criteria were made.

A. Effects of approach and environmental parameters on touchdown performance

The mixed-effects model analysis presented in Section IV.A was used to determine the experiment factors that had a significant effect on touchdown performance. This information provides insight into the most important parameters for a stabilized approach and the appropriate go-around gate height. The analysis revealed that touchdown performance was significantly more affected by the aircraft state at the 100-ft gate compared to the aircraft state at the 300-ft and 500-ft gates. The results suggest that at the 100-ft gate, the aircraft was too close to touchdown for the pilots to correct the unstable approach. The mean touchdown performance between the 300-ft and 500-ft gates did not vary significantly. The data indicate that the pilots were mostly able to correct the approach deviations at those gate heights. This confirms our previous findings that were based on data that did not include snapshots of the aircraft state at gates below the starting gate [5, 6].

The mixed model analysis was used to estimate the effect size of the approach parameters on touchdown performance. Localizer deviation had a significant effect on longitudinal touchdown point and lateral touchdown point at all gate heights. However, the plots in Fig. 9 show that the predicted longitudinal and lateral touchdown points are still within the desired touchdown box which suggests that the proposed criterion of less than one dot localizer deviation is acceptable from a touchdown performance perspective.

Glideslope deviation had a significant effect on sink rate at touchdown at all gate heights (see Table 10). GS deviation also had a significant effect on longitudinal touchdown point at the 500-ft and 100-ft gate heights. GS deviation had a significant effect on lateral touchdown point only at the 500-ft gate. Although the effect of GS deviation was often statistically significant, the second row of plot in Fig. 9 demonstrates that the predicted touchdown performance only exceeds the desired touchdown performance for a small subset of cases. For example, very high glideslope deviations (greater than 2 dots) for aircraft type 3 causes the predicted longitudinal touchdown point to be outside of the desired touchdown box. If glideslope deviation was less than a dot, the predicted touchdown performance was within the desired touchdown performance parameters; meaning that the proposed glideslope criterion should provide an acceptable level of safety.

Rate of descent only had a significant effect on touchdown dependent measures at the 100-ft and 500-ft gates. The third row of plots in Fig. 9 shows that ROD has a strong influence on both longitudinal touchdown point and sink rate at touchdown. There is little effect of rate of descent on any of the touchdown measures from the 300-ft gate; and, while there are statistically significant effects from the 500-ft gate, the plots show that the predicted touchdown performance changes little between different rates of descent at this gate height. Based on these findings, the recommendations of eliminating a ROD criterion from the go-around criteria seems to be acceptable from a touchdown performance perspective as long as the go-around decision occurs by the 300-ft gate.

At the 500-ft and 300-ft gates, airspeed had a significant effect on all touchdown dependent measures (see Table 10). As shown in the fourth row of Fig. 9, airspeed does have a strong effect on longitudinal touchdown point at the 100-ft gate, while the effect is less pronounced at the 300-ft and 500-ft gates. Again, this provides evidence that the go-around should occur by the 300-ft gate and that the proposed airspeed go-around criterion will likely result in the desired touchdown performance.

When the environmental variables of wind and visibility were introduced in Experiment 2, a significant effect on both the touchdown dependent measures and the subjective responses was observed. The mixed model analysis revealed that both wind and visibility had a significant effect on lateral touchdown point. Wind also had a significant effect on longitudinal touchdown point and sink rate at touchdown for most gate heights. In fact, during Experiment 2, wind had the strongest effect on touchdown performance, which signifies the importance of considering environmental parameters on approach and landing risk. However, it is important to note that excessive instabilities degrade landing performance even when visibility and wind conditions are more favorable.

B. Factors that Influence Go-Around Decision-Making

The analyses presented in the results section provided insight into the factors that drive go-around decision-making. The study revealed that pilots in all parts of the experiment base their go-around decision on a variety of factors including the approach parameters, environmental conditions, and their personal risk tolerance.

Of the approach parameters, the tablet questionnaire data revealed that high airspeed (V_{tgt} deviation) was a key driver in the go-around decision-making process. High airspeed was the primary reason that pilots selected when completing the question on why a go-around was performed, and the decision tree analysis performed on the tablet questionnaire results showed that the injected airspeed instability had the strongest effect on the outcome of each run in Experiment 3 and for A330 pilots in the first two experiments. Localizer deviation was found to be the second most important factor in the go-around decision-making process based on analyses of both the objective and subjective data. The decision tree analysis showed that the localizer deviation was the most important factor for runs in Experiment 3 where no airspeed instability was injected and for Boeing pilots in experiments 1 and 2. The results also showed that the glideslope deviation did not have as strong of an influence on the go-around decision, which might have been a consequence of the pilots flying using visual guidance as opposed to the instrument landing system (ILS).

One acknowledged shortcoming of the experiments is that the induced instabilities are highly unlikely to occur in real operations making several of the pilots uncomfortable and possibly leading to a higher number of go arounds. This is especially true for the induced localizer instability. However, pilots might also have been quicker with correcting the instabilities, as the nature of the instabilities became apparent after the first few runs, and pilots might have started to anticipate them, leading to fewer go arounds. This can be seen in the decision tree analyses where higher run numbers slightly decreased the probability of a go-around.

As shown by the decision tree analysis for experiments 1 and 2, visibility was one of the factors that influenced the go-around decision, especially for Boeing pilots when the localizer deviation was greater than 1 dot. This is likely because seeing the runway and getting back on the localizer is more difficult when the visibility has been reduced, and is perceived as more risky for the pilots, therefore visibility was not as strong of a factor for the go-around decision when there was no localizer deviation. The mixed model analysis on the dependent measures of perceived workload and risk also showed a significant effect of wind and visibility at both the 300 ft and 500 ft gates. The plots in Figs 13 and 14 show that as visibility decreased, perceived workload and risk tended to increase which in turn increases the likelihood that a pilot would prefer to go around. Similarly, runs with a crosswind rather than a tailwind tended to increase perceived workload and risk responses.

Generally, the pilots seemed to be more cautious when the environmental parameters were degraded and were more likely to go around in those situations. This pattern also seems to be true in the operational environment. That is, visual approaches can be associated with more pilot errors [15] and thus unstable approaches could be more often encountered and landed in visual conditions. Therefore, generally pilots are able to identify environmental parameters that increase risk and are more willing to go around during degraded environmental conditions. However, this also indicates that crews may become complacent when visibility and wind conditions are good, which could reduce the likelihood of a go around being performed and thus increasing the probability of landing an unstable approach. As the results of this study indicate, even with good visibility and no crosswind, touchdown performance can be degraded by approach parameters such as localizer deviation and excessive airspeed.

Based on the tablet questionnaire results, the decision to go around was solely based on the state of the aircraft and the environmental parameters on approach. Less than 1% of the respondents said their decision would have been different if the runway was longer, and no pilots said that their decision would have been different if the runway condition was better. The pilots in the study noted that this was because the acceptability of the runway had already been determined well before this stage of the approach.

C. The Effect of Pilot Variability on Dependent Measures and Questionnaire Responses

The results gathered strongly suggest that pilot variability has a significant influence on dependent measures. This was apparent while analyzing questionnaire responses, go around initiation heights, and go around decision making. For example, as shown in Fig. 12, the pilot rating of the risk of the scenarios varied widely. Although all the pilots in the study were exposed to the same set of scenarios, some responded with a low mean risk score of less than 5 out of 20, while others found many of the scenarios to be high risk and had mean risk scores of over 14 out of 20. This finding shows the pilots in the study had a varied amount of comfort level with the instabilities that were presented, which highlights the difficulty in developing stabilized approach criteria that will be broadly accepted.

While the linear mixed-effects model did show that pilots did contribute to the variability of the touchdown measures, the fixed effects in the experiment accounted for the majority of the variance. Of the three touchdown dependent measures, the pilot random effect explained the largest percentage of the response variance for the longitudinal touchdown point followed by sink rate at touchdown. This could indicate the difference in pilot preference for how to handle a high-energy approach (i.e., being high and fast). Some might try to extend the flare and land further down the runway, while others will opt to land in the touchdown zone with a high sink rate.

Because the pilots contributed to a relatively small amount of variability in touchdown performance dependent measures, this information can be used to make objective conclusions about the parameters and thresholds that should be used for go-around criteria. The high variability in the subjective responses by the pilots makes evaluating and refining the proposed criteria difficult using the subjective data; however, the data can still be used to provide some insight into the likelihood that a given set of criteria would be accepted by the pilots in the study.

D. Pilot Acceptance of Proposed Go-Around Gate and Criteria

The primary goal of Experiment 3 of the study was to determine whether pilots find the 300-ft go-around gate and the proposed go-around criteria acceptable. The tablet questionnaire data revealed that, generally, pilots rated the acceptability of the gate height and criteria high after each run; however, only 60% of the pilots stated the gate height and criteria were acceptable during the post-simulation written questionnaire at the end of the experiment. Note that the tablet questionnaire asked the acceptability of the criteria for that specific run, and the post-sim questionnaire for the acceptability in general. Many of the pilots that found the proposed gate height and criteria unacceptable in the post-sim questionnaire preferred a 500-ft or 1,000-ft go-around gate. Several of these pilots expressed concern about the risk of a 300-ft gate and did not believe there was a sufficient reason for lowering the gate height. The go-around decision point data supports this finding because more than half of the go arounds were executed prior to the 300-ft gate. However, several pilots stated that they were uncomfortable with the induced instabilities and chose to perform a go-around as soon as the aircraft became unstable, which might have skewed the data.

Although the tablet responses on the acceptability of the gate height and criteria were mostly favorable, analysis of the tablet questionnaire data showed that the pilots rated the acceptability of the criteria and gate height lower when airspeed was high. This could be a result of the compressed time-line before landing when the aircraft is fast. One takeaway from this outcome is that having checks of stability earlier in the approach is important to ensure that the aircraft is either within or trending to the stability criteria well before the 300-ft mark.

One measure of the validity of the proposed criteria was how often it was violated during the experiment (as discussed in [7]). In the B737, the criterion most often violated before landing was the airspeed limit. However, most were only 1 or 2 knots over the upper bound of the airspeed criterion. This highlights one of the difficulties in setting hard limits for stabilized approach criteria. Even though these flights were technically outside of the criteria, their landing performance was indistinguishable from the stable flights. For this reason, many pilots recommended that momentary deviations be permitted.

In the A330, most of the unstable flights that landed violated the glideslope criterion. One possible reason for this is that the pilots may have switched to a visual approach rather than using the ILS guidance. For this particular runway, the PAPI and ILS do not coincide, meaning that if the pilot was using the PAPI rather than the ILS, the glideslope could trend towards unstable. Additionally, in a wide-body aircraft, the ILS antenna is relatively far from the pilot's eyes, creating more discrepancy between ILS and visual guidance. This finding demonstrates the difficulty in defining hard limits for flight-path deviations. The several types of guidance that a pilot might be following to land the aircraft have to be considered; meaning either the criteria might have to be more complex to cover all types of flight path guidance, or flight path can only be loosely defined using such wording as "on flight path."

One question the pilots were asked after a run was landed was whether the aircraft was stable at 300 ft. This was to gauge whether pilots were able to correctly assess the approach relative to the proposed criteria. In most cases, based on the tablet responses, pilots correctly identified whether the aircraft was stable at the 300-ft gate (see [7]). When the pilots did misclassify the stability of the aircraft, they were typically either too fast or off the glideslope. It is possible that the turbulence made it difficult to determine whether the airspeed was stable. In many cases, the aircraft was only a couple knots fast, which is within the amount of airspeed fluctuation caused by moderate turbulence. These fluctuations might have led the pilot to believe the criteria were met when technically they were a little fast. The glideslope misclassification was likely affected by the PAPI and ILS not coinciding for the runway. Again, this highlights the difficulty of setting hard limits on approach parameters, and perhaps momentary deviations should be allowed.

In general, the acceptability of the criteria was highly dependent on each pilot's personal risk tolerance; therefore, it is unlikely that criteria can be developed that will satisfy every pilot's mental model of a stable approach. As the written questionnaire responses showed, pilots in the study were split on whether the proposed criteria and gate height were acceptable. The pilots who found the criteria unacceptable had a variety of reasons for their assessment, such as lack of an engine spooled parameter, lack of a rate of descent limit, and finding the airspeed criterion overly restrictive.

E. Final Proposed Criteria

The objective results of the study suggest that the proposed go-around criteria in Table 3 are acceptable for all three aircraft types studied. If the criteria were met at the 300-ft gate, the predicted touchdown performance is within the desired touchdown zone with an acceptable sink rate. The data also suggest that touchdown performance between equivalent aircraft states at the 300-ft and 500-ft gates result in similar touchdown performance outcomes. A unique feature of the proposed criteria compared to the traditional stabilized approach criteria is the elimination of the threshold for rate of descent and instead specifying that TAWS should not be activated. The objective findings of the study support the proposed criterion for two reasons:

1) Rate of descent at the 300-ft gate had little effect on touchdown dependent measures.

2) Momentary rate of descent deviations are common when correcting flight path which commonly causes an approach to be flagged as unstable; however, the study results suggest that if the momentary deviations are corrected by the 300-ft gate, the touchdown performance will not be compromised.

In addition, it should be noted that the TAWS activation is a function of altitude, such that the threshold rate of descent for TAWS activation is higher at higher altitudes and lower at lower altitudes. Therefore, the TAWS activation provides a better criterion and threshold when considering the relationship between a high rate of descent and the true safety risk of a given approach compared to a fixed rate of descent criterion for a go-around. For example, a given rate of descent at 1,000 ft AGL may be considerably less risky than the same rate of descent at 200 ft AGL, and the TAWS activation helps to differentiate the risk levels.

While the objective results suggest the proposed criteria are acceptable, the responses of several pilots in the study suggest that some modifications need to be made for broad acceptance among the pilot community. Although 60% of the pilots found the proposed criteria acceptable overall in the post-sim questionnaire, others cited a variety of changes that would need to be made in order to find the criteria acceptable.

The concern most often noted by the pilots was lowering the go-around gate to 300 ft. Several thought 300 ft was too low or did not believe there is sufficient evidence or reason to lower the gate. This concern may be mitigated by further emphasis on having stabilized approach gates at 1,000 ft and 500 ft with a final go-around gate at 300 ft (see Fig. 18), similar to [1]. Guidance should stress that a crew should only continue to the 300-ft gate if they believe that the criteria can realistically be met by that gate, and the crew should perform a go-around earlier than 300-ft if the crew is uncomfortable with the approach. The 300-ft gate should just serve as a final checkpoint to determine whether the approach should continue or if a go-around should be performed.

The proposed stabilized approach and go-around gates are to be used in both VMC and IMC conditions. Currently, many airline standard operating procedures have different stabilized approach gates depending on environmental conditions (500 ft for VMC and 1,000 ft for IMC). However, the final proposed criteria from this study are the same for both VMC and IMC conditions for a couple of reasons. First, the proposed criteria should be checked at the 1,000-ft, 500-ft, and 300-ft gates regardless of environmental conditions. This promotes situation awareness and helps the crew remain vigilant in any condition. Second, the objective results of the study suggest that if an approach egregiously violates the proposed criteria, the touchdown performance will be degraded regardless of the weather conditions. Wind condition had a significant effect on touchdown performance; however, note that the wind condition is accounted for by the criteria to some extent by the use of the target approach speed which includes a correction for wind.

The second most common aspect that pilots suggested changing about the proposed criteria was to add a requirement that engines must not be at an idle setting (i.e., engines must be spooled). This finding suggests that adding an engine spooled criterion could broaden the acceptance of the criteria among pilots. Ensuring the engines are spooled is also a good check to ensure that airspeed does not decay and rate of descent does not rapidly increase. Having the engines

Table 9 Final proposed go-around approach criteria.

	Criteria at 300 ft	
1	Airspeed	Within 0/+10 of target
2	Glideslope deviation (if available)	Less than 1 dot
3	Localizer deviation (if available)	Less than 1 dot
4	Rate of Descent	No TAWS activation
5	Thrust	Engines spooled

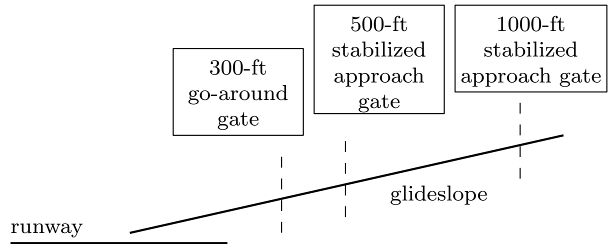

Fig. 18 Stabilized approach and go-around gates [1].

spooled also ensures that power is readily available in the event that a go-around needs to be performed close to the ground.

Therefore, based on this study, the go-around criteria in Table 9 are recommended. The criteria remained unchanged from the proposed criteria in Table 3, with the addition of an engine spooled parameter.

VI. Conclusions

Three experiments were conducted using B747-400, B737-800, and A330-200 Level D full-flight simulators with the objective to refine go-around criteria for transport category aircraft. A total of 30 crews or 60 pilots participated. In the first two experiments, focused on developing and validating new criteria from objective touchdown performance data, multiple approaches under different approach conditions and environmental variables were flown. Pilots were instructed to always land the aircraft, even from conditions considered to be an unstable approach. Various touchdown performance metrics were analyzed. In addition, pilots perceptions of risk under the various unstable approach conditions and resulting landings were assessed.

The third experiment captured pilot feedback and decision-making with regard to the developed go-around criteria from the first two experiment parts, and it assessed the crews' awareness of the aircraft state on approach. Pilots evaluated multiple approaches that were on the borderline of the developed go-around criteria when reaching 300 ft altitude. Pilots were instructed that they could either execute a go-around or land the airplane on each run, forcing a decision for the borderline cases at 300 ft. Pilots were instructed to go around if the aircraft was outside of the go-around criteria at 300 ft.

The following conclusions can be drawn based on the combined data analysis presented in this paper about universal go-around criteria for transport category aircraft:

1) Results show little difference in touchdown performance for conditions from the 300-ft and 500-ft gates. Conditions at the 100-ft gate introduced significant differences in touchdown performance, suggesting criteria using a 300-ft go-around gate.
2) Visibility and wind had little effect on touchdown performance compared to the approach parameters and were not dependent on the gate. This suggests that different criteria for IMC and VMC are not required. Wind

condition is accounted for by the criteria to some extent by the use of the target approach speed which includes a correction for wind.

3) In general, pilots rated the acceptability of the proposed criteria at the 300-ft go-around gate high at the end of each run. However, 40% of the pilots were uncomfortable with the 300-ft gate after completing the experiment. Placing more emphasis on checking approach stability and using active call-outs at 1,000 ft and 500 ft AGL might make more pilots comfortable using a 300-ft final go-around gate. Additional training may also be needed to reinforce the concept of having two stabilized approach gates and a go-around gate.
4) Allowing for momentary deviations from a stabilized approach should be considered, as long as they are corrected for by the 300-ft gate.
5) The subjective data suggest that there would be larger pilot acceptance of the proposed criteria if the criteria include an engine spooled parameter.

The objective and subjective results obtained from this study provided valuable information needed to develop updated go-around criteria. The proposed criteria performed well, and most pilots would find the criteria acceptable with some minor adjustments. The next step of this research is to conduct additional evaluation of the data collected including analyzing the simulation data from a time series perspective and further investigation of the relationship between the objective and subjective data collected. The researchers also plan to evaluate the operational implications of the proposed criteria through the study of operational data and collecting feedback at industry workshops. Furthermore, supplementary research is planned to study mitigations for potential risks when go-arounds are performed. With this approach, unstable approach rates might be reduced, and overall approach and go-around safety may be increased.

Appendix

This appendix provides more details on the mixed-effects models estimated to predict touchdown performance, and pilot workload and risk from the aircraft state (i.e. the independent variables) at the 500-ft, 300-ft, and 100-ft gates. Separate models were fit for the independent variables at each gate. The models included localizer deviation (ΔLOC), glideslope deviation (ΔGS), rate of descent (ROD), reference speed deviation (ΔV_{ref}), wind, visibility, and aircraft type as fixed effects. In addition, the interactions of ΔLOC, ΔGS, ROD, ΔV_{ref}, wind and visibility with aircraft type were included as fixed effects. Pilot role and the interaction of pilot role with aircraft were included as fixed effects for the workload and risk models only. Pilot was used as the random effect. The models included random intercepts only; that is, no random slopes were introduced.

The mixed-effects models were progressively built up by adding the different fixed effects and interactions one-by-one, starting with the intercept-only model. Table 10 provides the results of the likelihood ratio tests between models. The fixed factors were added to the model in the order from top to bottom in the table. Effects that were not significant were removed from the final model. As an example, the final mixed-effects model for the longitudinal touchdown point based on the independent variables at 500 ft is provided by (Table 10):

$$x_{td,i} = (\gamma_0 + u_{0i}) + \gamma_{1i}\Delta\mathrm{LOC}_i + \gamma_{2i}\Delta\mathrm{GS}_i + \gamma_{3i}\mathrm{ROD}_i + \gamma_{4i}\Delta V_{ref,i} + \gamma_{5i}\mathrm{aircraft}_i + \epsilon_i \quad (2)$$

with γ the mean model parameter estimates, u the random effect of pilot, and ϵ the residual term for the i-th pilot.

Acknowledgments

The authors would like to thank the simulation engineers and support personnel who contributed to the experiments. We especially thank Randall Cooper, Ricky Zoellner, and Larry Miller from the Flight Operations Simulation Laboratory at the FAA Mike Monroney Aeronautical Center in Oklahoma City for their valuable contributions in setting up and running the experiments. We also thank Doug Rodzon for being a test pilot on the A330 simulator. We would like to thank the FAA William J. Hughes Technical Center for the continued program support. Finally, we would like to thank the pilots who participated in the experiments and experiment-planning workshops.

Table 10 Mixed-effects Model comparison test statistics.

Independent Variables	Dependent Measures														
	x_{td}			y_{td}			$\dot{h}_{td}$			workload			risk		
	df	χ^2	p	df	χ^2	p	df	χ^2	p	df	χ^2	p	df	χ^2	p
500-ft Gate															
ΔLOC	1	15.33	<0.01**	1	73.21	<0.01**	1	3.51	0.06	1	52.27	<0.01**	1	186.62	<0.01**
ΔGS	1	19.08	<0.01**	1	28.78	<0.01**	1	13.98	<0.01**	1	11.10	<0.01**	1	71.96	<0.01**
ROD	1	7.92	<0.01**	1	11.24	<0.01**	1	6.54	0.01**	1	5.34	0.02**	1	9.52	<0.01**
ΔV_{ref}	1	13.42	<0.01**	1	6.35	0.01**	1	15.53	<0.01**	1	10.23	<0.01**	1	53.45	<0.01**
wind	2	2.81	0.24	2	267.42	<0.01**	2	20.92	<0.01**	2	25.93	<0.01**	2	14.41	<0.01**
visibility	1	0.24	0.63	1	13.90	<0.01**	1	0.87	0.35	1	77.89	<0.01**	1	191.75	<0.01**
aircraft	2	36.02	<0.01**	2	23.12	<0.01**	2	49.90	<0.01**	2	8.55	0.01**	2	7.76	0.02**
ΔLOC × aircraft	2	2.32	0.31	2	9.93	0.01**	3	4.24	0.24	2	6.96	0.03**	2	7.54	0.02**
ΔGS × aircraft	2	4.14	0.13	2	1.25	0.54	2	17.57	<0.01**	2	1.91	0.38	2	0.62	0.73
ROD × aircraft	2	4.43	0.11	2	0.64	0.72	2	0.89	0.64	2	0.19	0.91	2	0.71	0.70
ΔV_{ref} × aircraft	2	5.53	0.06	2	2.17	0.34	2	5.76	0.06	2	1.69	0.43	2	0.61	0.74
wind × aircraft	4	6.62	0.16	2	18.91	<0.01**	2	3.24	0.20	2	0.53	0.77	2	0.76	0.68
visibility × aircraft	2	5.01	0.08	1	6.27	0.01**	2	2.44	0.30	1	0.46	0.50	1	2.53	0.11
pilot role	–	–	–	–	–	–	–	–	–	1	281.77	<0.01**	1	23.22	<0.01**
pilot role × aircraft	–	–	–	–	–	–	–	–	–	2	11.73	<0.01**	2	2.39	0.30
300-ft Gate															
ΔLOC	1	23.64	<0.01**	1	103.44	<0.01**	1	2.28	0.13	1	70.05	<0.01**	1	152.37	<0.01**
ΔGS	1	0.52	0.47	1	0.96	0.33	1	12.08	<0.01**	1	0.37	0.54	1	1.90	0.17
ROD	1	1.11	0.29	1	1.28	0.26	1	1.11	0.29	1	2.59	0.11	1	15.24	<0.01**
ΔV_{ref}	1	6.12	0.01**	1	5.25	0.02**	1	15.33	<0.01**	1	4.43	0.04**	1	58.36	<0.01**
wind	2	46.58	<0.01**	2	491.81	<0.01**	2	58.40	<0.01**	2	28.18	<0.01**	2	43.62	<0.01**
visibility	1	7.16	0.01**	1	21.62	<0.01**	1	5.89	0.02**	1	27.89	<0.01**	1	82.99	<0.01**
aircraft	2	30.41	<0.01**	2	23.29	<0.01**	2	49.81	<0.01**	2	6.28	0.04**	2	5.41	0.07
ΔLOC × aircraft	2	5.08	0.08	2	2.62	0.27	3	1.48	0.69	2	11.90	<0.01**	2	32.46	<0.01**
ΔGS × aircraft	3	5.80	0.12	3	1.35	0.72	2	3.42	0.18	3	4.03	0.26	3	13.80	<0.01**
ROD × aircraft	3	7.09	0.07	3	4.22	0.24	3	2.74	0.43	3	7.68	0.05	2	6.27	0.04**
ΔV_{ref} × aircraft	2	15.30	<0.01**	2	0.16	0.92	2	4.13	0.13	2	3.67	0.16	2	1.82	0.40
wind × aircraft	2	3.22	0.20	2	35.68	<0.01**	2	3.51	0.17	2	1.54	0.46	3	5.28	0.15
visibility × aircraft	1	15.80	<0.01**	1	3.76	0.05	1	5.52	0.02**	1	3.61	0.06	1	5.25	0.02**
pilot role	–	–	–	–	–	–	–	–	–	1	203.46	<0.01**	1	23.22	<0.01**
pilot role × aircraft	–	–	–	–	–	–	–	–	–	2	1.51	0.47	2	0.80	0.67
100-ft Gate															
ΔLOC	1	26.74	<0.01**	1	182.57	<0.01**	1	5.52	0.02**	–	–	–	–	–	–
ΔGS	1	436.91	<0.01**	1	2.36	0.12	1	11.79	<0.01**	–	–	–	–	–	–
ROD	1	5.75	0.02**	1	0.09	0.76	1	39.26	<0.01**	1	4.44	0.04**	1	25.05	<0.01**
ΔV_{ref}	1	377.61	<0.01**	1	1.85	0.17	1	2.67	0.10	1	9.52	<0.01**	1	59.16	<0.01**
wind	2	51.50	<0.01**	2	400.35	<0.01**	2	52.28	<0.01**	–	–	–	–	–	–
visibility	1	0.32	0.57	1	16.46	<0.01**	1	2.15	0.14	–	–	–	–	–	–
aircraft	2	43.07	<0.01**	2	28.13	<0.01**	2	53.98	<0.01**	2	13.93	<0.01**	2	19.07	<0.01**
ΔLOC × aircraft	2	7.43	0.02**	2	6.70	0.04**	2	0.46	0.80	–	–	–	–	–	–
ΔGS × aircraft	2	8.52	0.01**	3	12.14	0.01**	2	6.10	0.05**	–	–	–	–	–	–
ROD × aircraft	2	41.09	<0.01**	3	1.19	0.76	2	3.65	0.16	2	1.13	0.57	2	4.96	0.08
ΔV_{ref} × aircraft	2	44.88	<0.01**	3	1.78	0.62	3	13.18	<0.01**	2	1.06	0.59	2	0.28	0.87
wind × aircraft	2	1.00	0.61	2	36.11	<0.01**	2	3.74	0.15	–	–	–	–	–	–
visibility × aircraft	2	7.94	0.02**	1	2.43	0.12	2	10.03	0.01**	–	–	–	–	–	–
pilot role	–	–	–	–	–	–	–	–	–	1	45.34	<0.01**	1	10.28	<0.01**
pilot role × aircraft	–	–	–	–	–	–	–	–	–	2	1.05	0.59	2	0.39	0.82

** = significant ($p < 0.05$)

– = not applicable

References

[1] Blajev, T., and Curtis, W., "Go-Around Decision-Making and Execution Project," Final report, Flight Safety Foundation, Mar. 2017. URL `https://flightsafety.org/wp-content/uploads/2017/03/Go-around-study_final.pdf`.

[2] International Air Transport Association, "Unstable Approaches, Risk Mitigation Policies, Procedures and Best Practices," Final report, 2017.

[3] Commercial Aviation Safety Team, "Runway Excursion - Flight Crew Landing Training," Safety Enhancement SE 216, Jun. 2014. URL `https://www.skybrary.aero/bookshelf/books/2806.pdf`.

[4] Federal Aviation Administration, "Mitigating the Risks of a Runway Overrun Upon Landing," , No. AC 91-79A, 2014.

[5] Campbell, A., Zaal, P., Schroeder, J. A., and Shah, S., "Development of Possible Go-Around Criteria for Transport Aircraft," *2018 Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, 2018. doi:https://doi.org/10.2514/6.2018-3198.

[6] Zaal, P., Campbell, A., Schroeder, J. A., and Shah, S., "Validation of Proposed Go-Around Criteria Under Various Environmental Conditions," *2019 Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, 2019. doi:https://doi.org/10.2514/6.2019-2993.

[7] Campbell, A., Zaal, P. M. T., Shah, S., and Schroeder, J. A., "Pilot Evaluation of Proposed Go-Around Criteria for Transport Aircraft," *2019 Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, 2019. doi:https://doi.org/10.2514/6.2019-3610.

[8] Sullivan, B., and Soukup, P., "The NASA 747-400 flight simulator - A national resource for aviation safety research," *Flight Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, 1996. doi:https://doi.org/10.2514/6.1996-3517.

[9] JMP®, Version 13, Cary, NC: SAS Institute, 1989-2007.

[10] MATLAB, "Version R2013b," The MathWorks Inc., 2013.

[11] SAS Institute Inc., "Using JMP® 13," SAS Institute Inc., Cary, NC, 2016.

[12] Box, G. E. P., and Cox, D. R., "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Vol. 26, No. 2, 1964, pp. 211–252.

[13] Shah, S., and Campbell, A., "Analyzing Pilot Decision-Making Using Predictive Modeling," *International Conference for Research in Air Transportation*, 2018.

[14] SAS Institute Inc., "JMP® 15 Predictive and Specialized Modeling," Tech. rep., 2016.

[15] Civil Air Navigation Services Organization, "Unstable Approaches ATC Considerations," Available at `https://www.icao.int/safety/RunwaySafety/Documents/`, 2011.