Loss of Control in Flight: comparing qualitative pilot opinion with quantitative flight data

Joao P. Macedo* and Jorge H. Bidinotto†
Department of Aeronautical Engineering, University of Sao Paulo, Joao Dagnone Av., Sao Carlos, 13568-250, Brazil

Michael A. Bromfield‡
Institute for Future Transport and Cities, Coventry University, Gulson Street, Coventry, CV1 2JH, United Kingdom

Loss of Control In Flight (LOC-I) is the most lethal type of accident in recent aviation history and the one that has experienced the least reduction in the past 20 years. These, combined with recent accidents such as Lion Air 610 and Ethiopian Flight 302, help explain the attention dedicated by several aviation stakeholders to prevent it from happening. Prior to that, however, the understanding of the characteristics of the phenomenon is of paramount importance. Pilots, aircraft and systems are deeply related in LOC-I accidents, but current definitions tend to examine these in isolation and not as a human-in-the-loop system. Following experiments in a flight simulator with a group of test pilots, this paper investigates the use of the Cooper-Harper rating scale and the Quantitative Loss of Control Criteria to assess controllability in a series of manoeuvres performed in four different sets of LOC-I test scenarios. The experimental design, including the selection of pilots, adaptation of test scenarios, conception of tasks and data gathering procedure, is thoroughly discussed. From a statistical correlation analysis, results indicate the existence of a positive weak to moderate monotonic relationship between the qualitative and quantitative approaches, however the resemblance of events indeed impacting on controllability issues is low. Finally, a discussion is taken on ways to conceive tasks with regard to the application of the Cooper-Harper scale and the possibility of adapting the quantitative method, both in the sense of being more faithful to the actual aircraft condition and pilots' workload.

Nomenclature

AA = Adverse Aerodynamics
$CFIT$ = Controlled Flight Into Terrain
CG = Centre of Gravity
DOF = Degree of Freedom
DPC = Dynamic Pitch Control
DRC = Dynamic Roll Control
GS = Glideslope
HRC = High-Risk Accident Occurrence Categories
HQR = Handling Qualities rating, in this case, Cooper-Harper rating
ILS = Instrument Landing System
IQR = Interquartile Range
KTS = knots
LDG = Landing Gear

---

*Corresponding Author, MSc Student, Department of Aeronautical Engineering, joao.macedo@usp.br, Brazil.

†Human Factors Researcher/Professor, Department of Aeronautical Engineering, Brazil.

‡Senior Lecturer in Aerospace, Institute for Future Transport and Cities, Faculty of Engineering and Computing, Coventry University, AIAA Member.

$LOC-I$ = Loss of Control in Flight
LOC = Localizer
MLW = Maximum Landing Weight (lbs)
n = Vertical load factor
N = Sample size
OEI = One Engine Inoperative
PFD = Primary Flight Display
PID = Proportional-Integral-Derivative
PIO = Pilot Induced Oscillation
QLC = Quantitative Loss of Control Criteria
r_S = Spearman correlation coefficient
SI = Structural Integrity
UA = Unusual Attitude
V_{FE} = Maximum operating equivalent airspeed for flaps up (kts)
V_{MO} = Maximum operating equivalent airspeed for flaps down (kts)
V_{REF} = Reference Speed (kts)
V_{SW} = Stall warning equivalent airspeed in 1-g flight (kts)

α = Angle of attack (deg)
α_{SW} = Angle of attack for stall-warning activation (deg)
β = Sideslip angle (deg)
β_{MDXW} = Sideslip angle for a non-crabbed approach in the maximum demonstrated crosswind for takeoff or landing (deg)
θ = Pitch angle (deg)
θ' = Dynamic pitch attitude (deg)
$\dot{\theta}$ = Pitch rate (deg per second)
ϕ = Bank angle (deg)
ϕ' = Dynamic roll attitude (deg)
$\dot{\phi}$ = Roll rate (deg per second)

I. Introduction

With increasing numbers of air passengers and competition, the challenge of maintaining safe and affordable air travel has never been greater. Based on the concept of HRC [1], two events highlighted: CFIT and LOC-I. The first consists of the collision of a completely under control aircraft with terrain, whilst the second is characterised by situations in which both the crew and the autoflight systems are incapable of controlling the aircraft flight path [2]. Despite being relatively uncommon, either type almost invariably results in deaths: from 2012 to 2016, 84% of CFIT occurrences were fatal, whilst 90% of LOC-I accidents yielded in casualties [3]. In the past two decades, however, CFIT experienced a significant reduction in its rate mainly due to the advent and diffusion of ground proximity warning systems [4], since the accident is typically associated with the approach phase (Fig. 1)*. On the other hand, LOC-I accidents permeate all phases of flight (Fig. 1) and the development of LOC-I-oriented prevention defences faces profound difficulties, so deep that the industry has not incorporated a widespread technology for the mitigation of this type of event [8].

Recent and highly publicised accidents, such as Lion Air 610 and Ethiopian Flight 302, show that the introduction of systems that inadvertently intervene in the human-machine interaction trying, for example, to avoid potential LOC-I precursors, like stall, may end up consolidating the accident instead of effectively helping. Mitigation of LOC-I events is currently a common aim amongst the major aviation stakeholders [3, 4, 9–12] and society urges for an issue softening, however LOC-I occurrences are highly complex and the latest events demonstrate that a clearer and more comprehensive understanding of these accidents is vital for the successful development and implementation of defences [13]. Being LOC-I, a dynamics and control problem involving the human-systems integration [2], it must be investigated from the

---

*Data are retrieved from IATA annual Safety Reports – [3, 5–7]

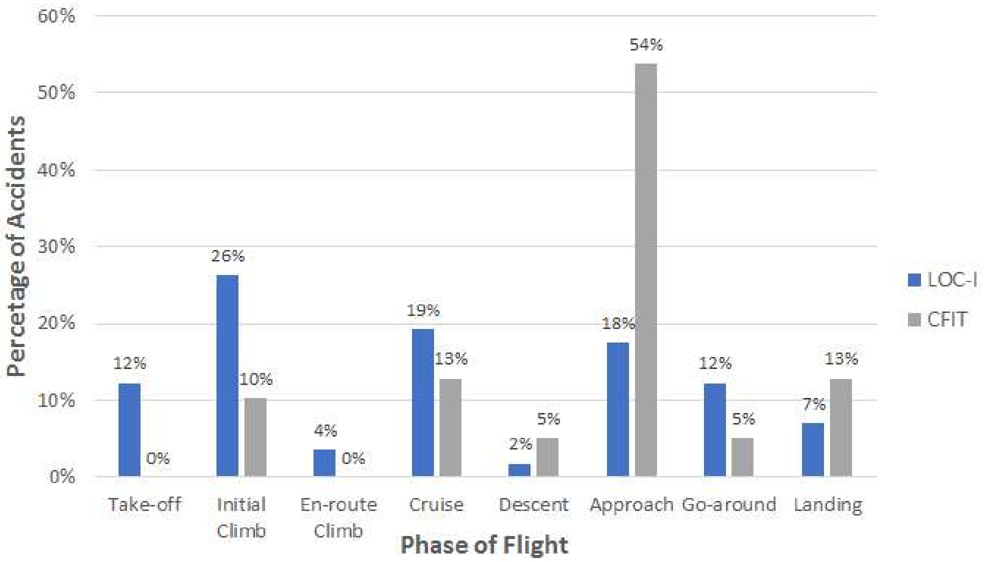


Fig. 1 CFIT and LOC-I percentage distribution per phase of flight (period: 2009 to 2016).

perspective of *pilots* and *aircraft*.

This paper seeks to examine the correlation between human perception and aircraft behaviour in potential LOC-I conditions. To assess the human point of view, the Cooper-Harper handling qualities rating scale is used, as its higher ratings foresee controllability threatening [14]; on the other hand, the Quantitative Loss of Control Criteria provides information in the aircraft perspective by the observation of key aircraft variables in LOC-I events [15]. The analysis is based on pilot-in-the-loop experiments conducted in a full-flight simulator with different human pilots. Section II details the Cooper-Harper scale and the QLC, meanwhile the simulator equipment, task definition, experiment setup and evaluation procedure are presented in Section III. Subsequently, qualitative and quantitative simulation results are analysed in Section IV and the correlation between these is discussed. Concluding remarks are finally provided.

II. Assessing *controllability*

Controllability is a matter of fundamental importance for the aeronautical industry simply due to the fact that, without control, aircraft are back to the beginning of the 20th in which ingenious solutions were proposed and tested to make it humanly flyable, unfortunately, sometimes with the worst consequences. Nowadays, after years of development in aviation and with increasing levels of automation within the cockpit, controllability is still vogue, but now especially focused in the perspective of the human-aircraft interaction. As difficult as it can be, the assessment of *controllability* is vital not only for aircraft certification purposes, but also to understand LOC-I events and promote safety enhancements. Due to the characteristics of LOC-I occurrences, independent mechanisms used to derive information about aircraft controllability are selected: (i) the Cooper-Harper Handling Quality Rating Scale, to address the human being qualitative point of view, and (ii) the Quantitative Loss of Control Criteria, geared towards the aircraft itself and representing a quantitative approach. For what follows, these are detailed.

A. The Cooper-Harper rating scale

Controllability, in the human pilot viewpoint, may be understood as the capacity to maintain/change the aircraft attitude according to his/her intentions and inputs; the characteristics that govern the ease and precision in which control can be exerted are named *handling qualities* [14]. Naturally, therefore, *control* and *handling qualities* are intimately linked and the mechanisms used to evaluate the handling qualities of an aircraft go through the observation of the vehicle's control features, being that exactly what the Cooper-Harper scale does.

Essentially, the mechanism is a 10-point scale arranged in the form of dichotomous decisions whose ratings reflect the human pilot evaluation about the aircraft handling qualities based on the binomial *performance-workload* in the context of a given task . *Performance* refers to the precision of aircraft control, being it separated in two levels (to wit, *adequate* and *desired*, meanwhile *workload* consists of the measure of how much effort, mental and physical, is the pilot supplying to attain a given level of performance within the assigned task. In fact, *performance* and *workload* are so strongly linked that one cannot be characterised independently of the other, therefore, a separate variable, named *compensation*, is used to to conceive a proper pilot assessment [14].

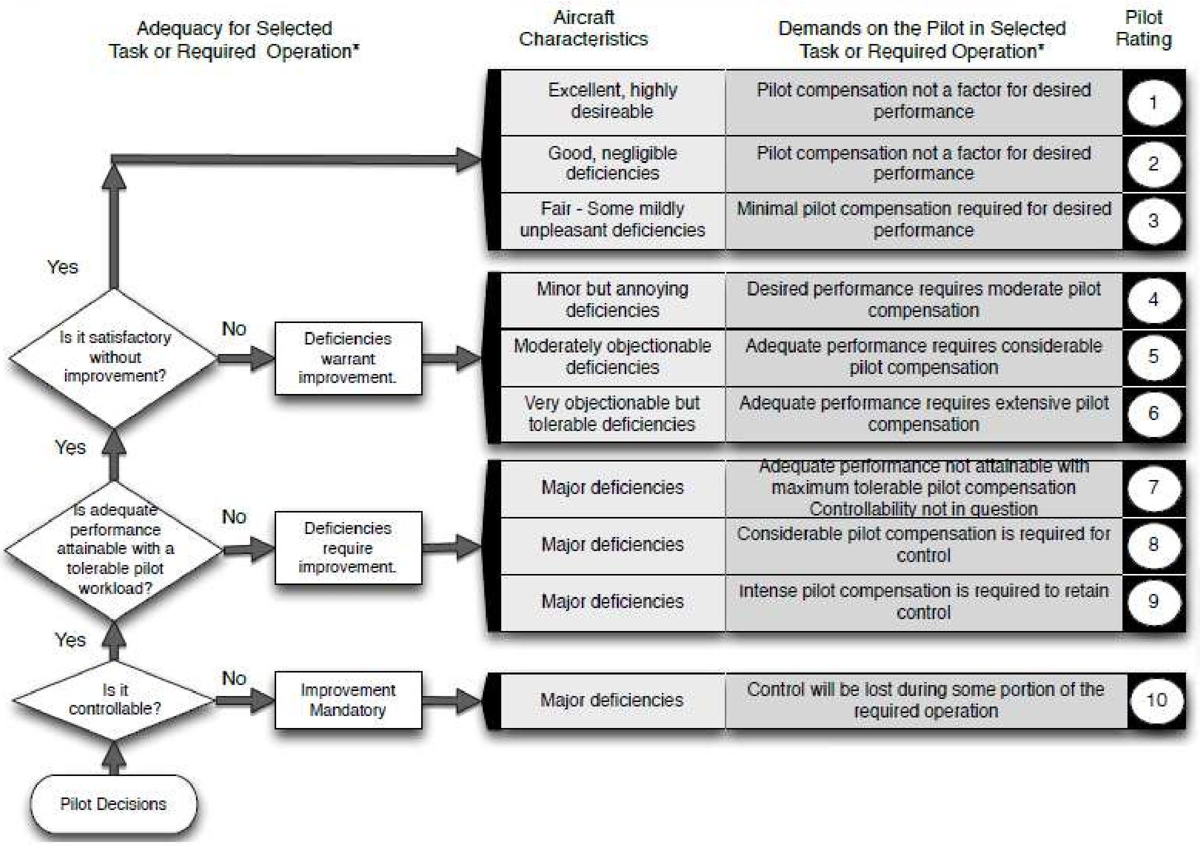


Fig. 2 The Cooper-Harper rating scale – adapted from [14].

Compensation intends to measure the **additional** pilot effort necessary to maintain a certain level of performance in consideration of deficient aircraft characteristics, hence, the total workload to perform a given task is comprised of the compensation **plus** the workload due to task itself [14, 16]. This way, pilots rate *handling qualities* looking at *performance* and *compensation*, however, behind the mask, the scale translates itself into *performance* and *workload*, the original binomial of interest. More specifically, the scale foresees two possible levels of performance, *adequate* and *desired*, and several degrees of compensation, from "not a factor" to "maximum tolerable", thus, to the use of the scale, for a given vehicle condition, pilots are assigned a certain task, with expected levels of performance previously determined and, based on their perceived workload and the achieved performance during the execution of the manoeuvre, rate the aircraft *handling qualities*.

Figure 2 presents the scale. In a wider look, it is possible to see **(a)** its decision tree structure, with the inquiries being on the left side and the numerical ratings displayed to the right, coming accompanied by a textual description, and **(b)** that ratings are grouped into four categories (or, as commonly mentioned, four levels). In a more detailed look, ratings *1* to *7* clearly show the correspondence between that given numerical rating with the pair *performance-compensation*, beyond that, ratings *8* to *10* are more delicate since they openly expose the dependence between handling qualities and controllability; for these ratings, performance is not a factor anymore as the control of the aircraft for the task under evaluation may be threatened, consequently making pilots turn their full attention to control the vehicle and not anymore to reach certain performance.

In a controllability perspective, therefore, the numerical ratings may be regrouped to conceive three distinguished categories: (i) control is *not* a problem, corresponding to Cooper-Harper ratings varying from *1* to *7*; (ii) control is threatened, being that a *near*-LOC-I condition, represented by HQRs *8* and *9* and, finally, (iii) situations in which control is lost or, in other words, a LOC-I condition, attested by the HQR *10*. This classification is of primary relevance for the analysis of the results obtained.

It is still important to mention that, despite the ease of having the assessment of handling qualities translated into numerical ratings through the scale, indeed, the numbers are actually just a portion of the qualitative evaluation and be based solely on the numerical information oversimplifies the assessment. In fact, numbers do not provide details on *how* and *why* the evaluation pilot arrive at that classification, therefore the aviators should be stimulated to provide their objections and feelings, which may be done by a sort of "open-questionnaire", also named a comment card, consisting of a list of items for which comments are desired and that pilots should address for every task performed [16]. Details on the use of the Cooper-Harper scale during the experiments are provided in Section III.E.

B. The Quantitative Loss of Control Criteria

The Quantitative Loss of Control Criteria is currently the only mechanism based on quantitative variables to classify an event as a LOC-I. It consists of a parametric analysis in which ten aircraft variables, identified as the most important parameters for these events, are plotted in five envelopes relating to flight dynamics characteristics, flight control use (both for pitch and roll), structural integrity and aerodynamics. Figure 3 presents the variables used to conceive the criteria and the corresponding envelopes, which are briefly presented (details are found in [15]):

- **Adverse Aerodynamics – AA – envelope**: it plots the normalised angle of attack (α_{NORM}) versus normalised sideslip angle (β_{NORM}). The normalisation of α depends on the angle of attack that activates the stall warning system (α_{SW}), meanwhile β is normalised by the sideslip angle for a non-crabbed approach in the maximum demonstrated crosswind for takeoff or landing (β_{MDXW}). The AA envelope indicates stall conditions both in pitch and yaw;
- **Unusual Attitude – UA – envelope**: it maps the pitch attitude angle (θ) versus the bank angle (ϕ), thus plotting data that crew rely on most, as pilots are constantly looking at these flightpath parameters, once they are shown in the *attitude indicator*. The axes limits[†] are $-45° \leq \phi \leq +45°$ and $-10° \leq \theta \leq +25°$;
- **Structural Integrity – SI – envelope**: it is function of the vertical load factor (n) and normalised airspeed (V_{NORM}), indicating overspeed occurrence and structural overload. Both axes are dependent on the configuration of the aircraft, as shown in Fig. 3; the normalisation of the airspeed depends on the stall warning equivalent airspeed in 1-g flight (V_{SW}) and the maximum operating equivalent airspeed for flaps-up (V_{MO}) – Equation 1 – or flaps-down (V_{FE}) – Equation 2.

$$V_{NORM} = \frac{V_E - V_{SW}}{V_{MO} - V_{SW}} \qquad (1) \qquad\qquad V_{NORM} = \frac{V_E - V_{SW}}{V_{FE} - V_{SW}} \qquad (2)$$

- **Dynamic Pitch Control – DPC – envelope**: it plots dynamic pitch attitude (θ’) versus pitch control percentage usage. Dynamic pitch attitude (θ') is the sum of the current pitch with the pitch rate for one second – Equation 3 –, thus the envelope exposes consistency or opposition between pitch control commands and longitudinal aircraft motion;

[†] Given the ongoing discussion by aviation authorities regarding the definition of the term *usual attitude* [17, 18], the metrics herein adopted consider the values the scope proposed in [15]

$$\theta' = \theta + \dot{\theta} \qquad (3) \qquad\qquad \phi' = \phi + \dot{\phi} \qquad (4)$$

- **Dynamic Roll Control – DRC – envelope**: it is the equivalent of the DPC applied for the lateral axis, hence the dynamic roll attitude (ϕ') is the sum of the current bank angle with the roll rate for one second – Equation 4 –. The envelope provides information whether the utilisation of roll control commands is consistent or opposed to the lateral aircraft motion.

The bold lines in each envelope shown in Fig. 3, therefore, refers to the expected limits of that given variable in normal flight conditions, including emergency situations foreseen in manuals; the quantitative criteria is based on the number of envelopes crossed, i.e.: "normal operational maneuvers, even if aggressive, usually do not exceed more than one envelope[,] a maneuver that exceeds only two envelopes is a borderline LOC-I condition [and] a maneuver that exceeds three or more QLC envelopes can be classified as LOC-I [since it] seems to be a good working definition" [15]. It is possible to observe the number of envelopes crossed according to the same categories in which the HQRs have been rearranged, that is: (i) control is *not* an issue for manoeuvres in which zero or one envelope is crossed; (ii) control is threatened, a *near*-LOC-I condition, for tasks corresponding to the extrapolation of two QLC envelopes and, finally, (iii) control is lost, a LOC-I, when three, four or five envelopes are excursed. These groups are of major importance for the discussion of the results obtained.

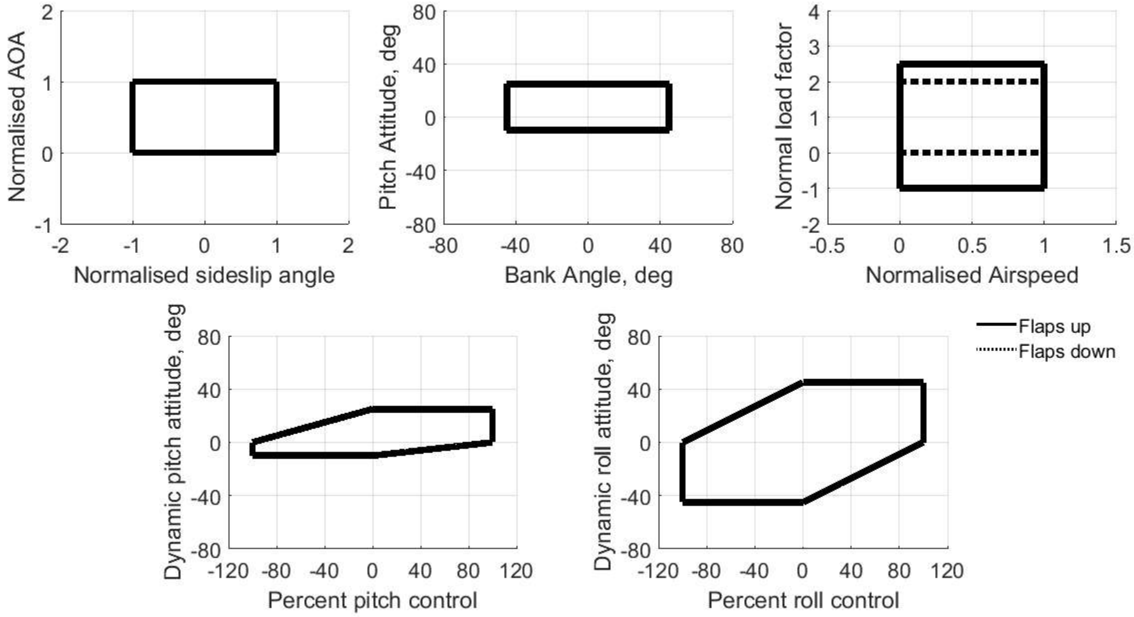


Fig. 3 Parametric envelopes – adapted from [15].

III. Experimental setup and procedures

Pilot-in-the-loop simulations were performed based on procedures previously adopted in human factors experiments and special attention was dedicated to the proper use of the Cooper-Harper rating scale aiming to reduce pilot variability in the qualitative assessment. The number of pilots and test points were shortened due to time restrictions and also pilot availability, given that experienced test pilots in flight test engineering evaluations are rare. Even though a thorough statistical analysis requires more data, [19] points that six different pilots represent a good balance point between cost and quality of the results in terms of confidence interval, the exact sample size of the experiments. In the sequence, aspects relating to the experimental setup and adopted procedures are detailed.

A. Apparatus

Simulations were performed on the 6DOF – Degree of Freedom – research simulator at the Sao Carlos School of Engineering (University of Sao Paulo | EESC-USP). Inside it, the only existing seat – normally occupied by the pilot – is surrounded by commercial controls and screens (Fig. 4b): a sidestick on the left-hand side, together with an armrest, a pedestal (with engine throttle and flap levers, among other configurable switches) on the right side and the rudder pedals in front of the seat, slightly downwards; a 24 inch screen emulating the aircraft panel with all necessary information is placed on a position equivalent to a head-down display and, finally, at sight of the pilot's eyes, a 50 inch screen displays the out-the window visual scene (Fig. 4a); a black cover "wraps" the cockpit to avoid external influence during the simulations (Fig. 4c). Flight deck instrumentation, out-the-window view and aircraft dynamic models are simulated through the programmable open-source software *FlightGear*, meanwhile the Stewart platform motion system is electrically actuated and controlled by a PID – Proportional-Integral-Derivative – and a washout filter programmed in the MATLAB/Simulink® environment [20]. For safety reasons, saturation limits are incorporated both to the control algorithm and the electrical actuators, yielding in the operational system motion capabilities provided in Table 1 (relative to the centre position of the platform).

Table 1 Platform saturation limits.

Degrees of freedom	Displacement
Longitudinal	-370 mm ≤ x ≤ 400 mm
Lateral	-400 mm ≤ y ≤ 400 mm
Vertical	-275 mm ≤ z ≤ 275 mm
Roll	$-15° \leq \phi \leq 15°$
Pitch	$-15° \leq \theta \leq 15°$
Yaw	$-15° \leq \psi \leq 15°$

B. Participants

Six different pilots, most of them currently employed by Embraer S/A, were selected to participate in the test campaign; Table 2 summarises their individual experience and background. Despite the relatively small average flight hours of the invited aviators, it is important to mention that the flight test engineering activity is much different than regular commercial flights, therefore, this "small" sum of flight hours in fact shades a great experience in a variety of handling qualities conditions, a pretty much desirable characteristic to reduce pilot variability regarding the qualitative ratings [19], moreover, pilots were all experienced to assess handling qualities using the Cooper-Harper rating scale.

Table 2 Evaluation pilots in the experiment campaign.

Pilot	Flight Hours	Type Ratings
1	19000	F100; B737-200/-300/-400; MD-11; A310; A319; A320; A330; A350
2	7500	E110; E120; E145; E170; E190; E550; L500
3	6000	E110; E120; E145; E170; E190; E550; L500
4	7500	AMX; E110; E120; E145; E170; E190; L500
5	7500	AT-26; C-95; E110; E120; E145; E170 E190; F-5; HS-125; L500
6	5800	E50P; E55P; E120; E145; E170; E190; E550

(a) Cockpit view | embarked screens.

(b) Cockpit view | installed controls.

(c) Outside view.

Fig. 4 EESC-USP's flight simulator.

C. Test scenarios

Air crash accidents, and even the departure to an "outside the flight envelope condition", do not happen from single causes, rather, they are the product of a particular combination of many contributing factors of different orders [21, 22]. Based on real LOC-I accidents, [23] identified their precursors and worst-case combinations to conceive a total of sixty LOC-I test scenarios intended to be used for research and even mitigation defences certification purposes. Considering time frame restrictions, simulator capabilities and statistical significance, four test scenarios were selected and adapted for the experiments[‡]:

- **Scenario 01:** this set essentially consists of a vehicle impairment precursor in the form of a single-engine failure. Given that pilots are usually trained for OEI operations, their actuation does not figure among the contribution factors, hence, to the simulation of this scenario, engine number 2 is simply switched off (scenario *3* from [23]);
- **Scenario 02:** this conjunction combines precursors of two different natures, a vehicle impairment and an inappropriate crew response. The first comes in the form of a 75% elevator effectiveness loss, which in real-life would be naturally followed by the application of high-gain inputs, therefore, the second precursor is simulated as exacerbated crew actuation on controls. In terms of simulation procedures, the inappropriate pilots response factor is addressed by incorporating gains (greater than 1) to the normal control response curves, this way conceiving the exacerbated curves; Fig. 5 presents a comparison between normal and modified shapes[§] (scenario *2* from [23] –

[‡] The selected scenarios cover approximately 22% of a set of 126 real LOC-I occurrences and 16% of the 6087 casualties

[§] The degree of change from the normal to the exacerbated curve is, in fact, arbitrary, although, it is a sufficient method to address the representation of over-controlling inputs

Table 3 Summary of the simulation scenarios.

Scenario number	Summary	Precursors	
01	Single engine failure	100% thrust reduction in engine number 2	
02	Loss of control surface effectiveness together with exacerbated crew inputs	Loss of 75% of the elevator effectiveness	Exacerbated crew control inputs
03	Unresponsive engine together with exacerbated and delayed crew inputs	Engine number 2 unresponsive and locked at 50% of the total available thrust	Exacerbated and delayed crew control inputs
04	Icing impairment together with delayed crew inputs	Icing accumulation: sooner stall, control surfaces less effective, thrust asymmetry	Delayed crew control inputs

adapted);

- **Scenario 03:** this combination of precursors addresses conditions in which, again, vehicle impairments combine with inappropriate crew actuation. In the present case, the former is represented by an unresponsive engine to crew inputs, meanwhile the latter comes in the form of exacerbated and delayed crew inputs. For simulation purposes, engine 2 is locked in the 50% available thrust, and the sidestick control response curve is modified according to Fig. 5, as in Scenario 02. Finally, delayed inputs are addressed by reducing the sampling frequency of the sidestick, thus a certain pilot input signal lingers to effectively enter the aircraft dynamics (scenarios *20* and *34* from [23] – adapted);
- **Scenario 04:** this set foresees icing accumulation together with inappropriate crew actuation in the form of delayed inputs. In terms of simulation, one more time, a delay between pilot control actuation and aircraft response is provided by reducing the sampling frequency of the sidestick, on the other hand, in which concerns ice accretion, the aircraft model configuration file is altered to, comparatively to the original model, (i) reduce the wing and horizontal stabiliser angles of attack corresponding to the maximum lift coefficient by 25%, (ii) reduce ailerons deflection lift increment by 20%, meanwhile drag increment caused by the same deflection is increased of 40%, (iii) reduce elevator lift increment by 15% and augment drag variation by 40%, **(iv)** reduce engine number 2 maximum available thrust by approximately 45% and delay the engine response by a factor of five. Therefore, in the aggregate, the control surfaces of the "icing" aircraft are less effective than in normal cases, as less lift and more drag increments are produced once they are deflected, moreover, thrust asymmetry occurs, both because of differential maximum thrust and response time, and a stall may more easily happen[¶] (scenario *25* from [23] – adapted)

Table 3 summarises the selected test scenarios, presenting their key parameters in terms of addressed precursors.

D. Tasks and levels of performance

In order to establish even conditions for all pilots and consistent use of the Cooper-Harper rating scale, a set of tasks was conceived to evaluate possible control issues in the selected LOC-I test scenarios. In the understanding that controllability basically lays on (i) the capability of keeping a desired attitude and *(ii)* the possibility of changing the current attitude in accordance with pilot inputs, the manoeuvres were designed aiming to evaluate these competences being based on real-world demands, thus meaning that these are manoeuvres indeed possibly to be performed in a normal operational environment. Furthermore, control is *not* suddenly lost, on the contrary, it degrades in all of the

---

[¶] The degree of change on the "normal" aircraft dynamics to conceive the "icing" model is arbitrary, however, the modifications are completely in accordance with the proposals of [23] and are characteristic of ice accretion events

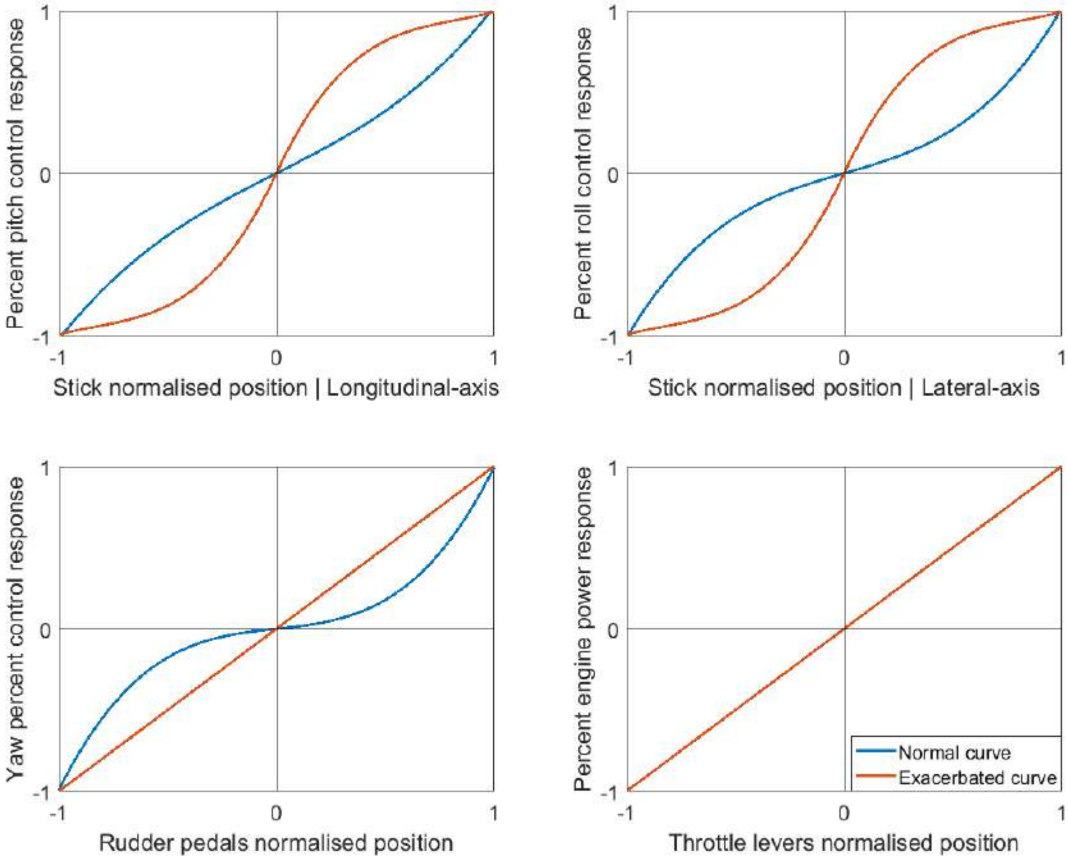


Fig. 5 Normal and exacerbated control response curves (stick & rudder + throttle inputs).

three axes, thus, some tasks are focused on the evaluation of controllability around the longitudinal axis, meanwhile some are more lateral-directionally demanding, still, a long-exposition manoeuvre combining inputs in all axes is also forecast. The tasks are briefly described in Table 5 together with their associated parameters under evaluation and *levels of performance*[‖].

Finally, due to time limitation and human factor reasons, not every task was performed in every scenario, this way, a greater number of pilots could participate in the research and results were less influenced by fatigue. Table 4 details the manoeuvres according to the set of precursors and also shows that, in total, 18 test points are under analysis.

Table 4 Tasks performed in each scenario.

Tasks	**1a**	**1b**	**2a**	**2b**	**3a**	**3b**	**4a**	**4b**	**7b**
Sc. 01	×	×	×	✓	✓	×	×	✓	×
Sc. 02	✓	×	✓	×	✓	×	✓	×	✓
Sc. 03	×	×	×	×	×	×	×	×	✓
Sc. 04	✓	✓	✓	✓	✓	✓	✓	✓	✓

[‖] Tasks *4a* and *4b* ask for bank angles not commonly indicated on the PFD, therefore, marks were provided to make possible the execution of these manoeuvres

Table 5 Collection of tasks and associated performance criteria.

Task ID	Objective	Description	Lateral-directional performance		Longitudinal performance		Timing performance	
			Desired	**Adequate**	**Desired**	**Adequate**	**Desired**	**Adequate**
1a	Maintain longitudinal attitude (climb)	Reach θ = 12.5° and maintain it until completing a 3000 ft climb. Maintain heading and airspeed within the required performance bands. No over- or undershoots.	Heading: ± 4°	Heading: ± 8°	θ: ± 2° Speed: ± 5 kts	θ: ± 4° Speed: ± 10 kts	—	—
1b	Maintain longitudinal attitude (descent)	Reach θ = -5° and attain it until completing a 3000 ft descent. Maintain heading and airspeed within the required performance bands. No over- or undershoots.	Heading: ± 4°	Heading: ± 8°	θ: ± 2° Speed: ± 5 kts	θ: ± 4° Speed: ± 10 kts	—	—
2a	Maintain lateral attitude (right turn)	Reach ϕ = 30° and attain it until completing a 90° turn (cross it with 30° bank). Maintain altitude and airspeed within the required performance bands. No over- or undershoots.	ϕ: ± 3°	ϕ: ± 6°	Altitude: ± 100 ft Speed: ± 5 kts	Altitude: ± 200 ft Speed: ± 10 kts	—	—
2b	Maintain lateral attitude (left turn)	Reach ϕ = -30° and attain it until completing a 90° turn (cross it with -30° bank). Maintain altitude and airspeed within the required performance bands. No over- or undershoots.	ϕ: ± 3°	ϕ: ± 6°	Altitude: ± 100 ft Speed: ± 5 kts	Altitude: ± 200 ft Speed: ± 10 kts	—	—
3a	Change longitudinal attitude (climb)	Reach θ = 12.5° as quickly and precisely as possible. Maintain heading and airspeed within the required performance bands. No over- or undershoots.	Heading: ± 4°	Heading: ± 8°	θ: ± 2° Speed: ± 5 kts	θ: ± 4° Speed: ± 10 kts	$\Delta t \leq 3$ sec	$\Delta t \leq 6$ sec
3b	Change longitudinal attitude (descent)	Reach θ = -5° as quickly and precisely as possible. Maintain heading and airspeed within the required performance bands. No over- or undershoots.	Heading: ± 4°	Heading: ± 8°	θ: ± 1.5° Speed: ± 8 kts	θ: ± 3° Speed: ± 15 kts	$\Delta t \leq 3$ sec	$\Delta t \leq 6$ sec
4a	Change lateral attitude (right turn)	Reach ϕ = 37.5° as quickly and precisely as possible. Maintain altitude and airspeed within the required performance bands. No over- or undershoots.	ϕ: ± 3°	ϕ: ± 6°	Altitude: ± 100 ft Speed: ± 5 kts	Altitude: ± 200 ft Speed: ± 10 kts	$\Delta t \leq 5$ sec	$\Delta t \leq 8$ sec
4b	Change lateral attitude (left turn)	Reach ϕ = -37.5° as quickly and precisely as possible. Maintain altitude and airspeed within the required performance bands. No over- or undershoots.	ϕ: ± 3°	ϕ: ± 6°	Altitude: ± 100 ft Speed: ± 5 kts	Altitude: ± 200 ft Speed: ± 10 kts	$\Delta t \leq 5$ sec	$\Delta t \leq 8$ sec
7b	Perform an ILS approach	Intercept and follow glideslope and localizer from an altitude of 2500 ft up to 500 ft AGL. Maintain altitude and airspeed within the required performance bands.	LOC offset: ± 1 *dot* No PIO	LOC offset: ± 2 *dots* No PIO	GS offset: ± 1 *dot* Speed: ± 8 kts No PIO	GS offset: ± 2 *dots* Speed: ± 15 kts No PIO	—	—

E. Evaluation procedure

Lasting approximately a total of four to five hours including resting stops, the simulations were all conducted with one pilot at a time and generally consisted of (i) preflight briefing; (ii) execution of the tasks and (iii) manoeuvre debriefing and Cooper-Harper rating assignment section. For what follows, details are provided on these three separate parts.

1. Preflight briefing

Beyond a technical section, the briefing also presented a psychological component fundamental to the success of the simulations and to the establishment of mutual confidence between the engineer in charge of the test campaign and pilots. Conducted in the form of a conversation, this section essentially revolved around the explicit description of the mission, translated as (i) what pilots are required to accomplish and (ii) the conditions under which the mission is conducted [14]. This way, pilots were presented in details to the nine predicted tasks to make clear their objectives, starting and ending points, the manoeuvres themselves, parameters under observation and tolerances defining the performance criteria predicted by the Cooper-Harper scale.

Relatively to the conditions of the mission (item (ii)), some depend on the simulation scenarios and tasks, meanwhile some are constant. Those constant, basically consist of the meteorological circumstances of the tests – daylight, no clouds, no wind, no turbulence and no traffic –, the selected aircraft – a Boeing 777-200ER in direct law control mode configured for MLW (Maximum Landing Weight) rear CG – and the manual pilot-in-the-loop control flight condition, without the aid of autopilot systems, as well as the necessity of performing the tasks based on instrument indications (instead of visual), because of the characteristics of the manoeuvres.

In order to address the conditions dependent on the simulation scenarios, a thorough discussion was established with pilots to determine the predicted behaviour of the aircraft under every combination of precursors and prepare the crewmember for the execution of the tasks, hence preventing eventual distractions caused by unknown system characteristics. Moreover, as mentioned by [24], the discussion was also important to make pilots clear that ratings are to be assigned to the aircraft "as is", i.e., to consider that the vehicle they are flying is ready to be produced and delivered to costumers, thus mitigating pilots tendency to take aircraft degradation into account and assign low HQRs because they know the vehicle is faulty and that mitigation defences in practice exist.

Concerning dependence on tasks, pilots were briefed about the aircraft configuration in each task, that is, every manoeuvre was performed with flaps retracted and gear up, exception made for task *7b* in which the aircraft was "dirty", meaning that flaps were in the fully extended position and the landing gear locked down. Furthermore, an identical set of the controls installed within the simulator cockpit was used to demonstrate the configuration of the switches, as well as to familiarise aviators with the equipment. Finally, the briefing section also served to detail the Cooper-Harper rating scale with the evaluation pilots, despite their familiarity, and clearly define the wording used on it to establish an equal level of understanding between aviators and diminish the impact of pilot variability in the results.

2. Execution of the tasks

Prior to execution of the tasks, pilots were allowed to perform a familiarisation flight with the aircraft fully operational to become more intimate to the simulator and aircraft behaviour; solely after stating to be comfortable, the experiments themselves actually began. For every scenario, the engineer configured the simulation software for the desired combination of LOC-I precursors and positioned the aircraft in the air, signalling to the pilot to assume the controls. For what followed, the aviator was the only responsible for the control of the aircraft, being able to familiarise her/himself with the condition of the vehicle in that given simulation scenario. When judging to be ready to commence the execution of the tasks, the aviator was instructed to fly and stabilise the aircraft in the condition (altitude and airspeed) corresponding to the starting point of the first manoeuvre, meanwhile s/he was remembered about what the task consisted of. When reaching and feeling comfortable in the required condition, the aviator signalled to the engineer her/his readiness to begin the task (a "TOP" mark), performed the manoeuvre itself and then, in a combined decision between engineer and pilot, stated another "TOP" mark to indicate the end of the task; pilots were still free to decide whether they wanted or not to repeat the point. During the execution of the task (between "TOP" marks), the engineer triggered a MATLAB® protocol to collect and record aircraft quantitative data necessary to the Cooper-Harper classification and to the QLC. In the hypothesis of pilots being satisfied with the execution of the task, the simulator was frozen to initiate the debriefing of manoeuvre, which culminates in the assignment of a HQR.

After debriefing the task, when the test pilot stated to be ready to proceed to another manoeuvre, the simulator was unfrozen and s/he was instructed about the next exercise and asked to fly to the new starting point, following the just mentioned procedure to execute the duty. Thus, in-between tasks of the same scenario, pilots were required to fly the vehicle, which points towards the *long-look* evaluation technique, another important aspect to minimise intrapilot variability [16]. In the occasion of the completion of all tasks of a certain simulation scenario, an identical procedure was applied to the next set of contributing factors, i.e., the engineer configured the simulator for the new scenario, placed the aircraft in the air and then asked the pilot to assume the controls for the whole sequence.

3. Manoeuvre debriefing and HQR assignment

Just after the completion of a certain task, being the simulator paused, a conversation was established between the evaluation pilot and the engineer to discuss the manoeuvre just performed. The talk was essentially based on the comment card conceived for the experiments (refer to Appendix A) and aimed to give pilots the opportunity to rationalise about the characteristics encountered during the exercise, a preparation for the Cooper-Harper scale, and the engineer the chance to better comprehend the human perception about the handling qualities of the aircraft and possible controllability issues.

Following the comment section, pilots were invited to use the Cooper-Harper scale to give a rating for the task. However, in turn, it meant that the engineer guided the test pilot within the assessment mechanism, i.e., the questions foreseen in the scale were accordingly posed by the engineer to the pilot, which, in conjunction with the level of performance information, could assign a final rating. The stiff procedure was indeed necessary and guaranteed the utilisation of the decision tree structure of the scale (if unattended, one of the reasons for scattered ratings), as well as conceived a thorough handling quality evaluation, since the other "half" of the scale was assured, pilots' comments.

Following the just described procedure, both qualitative and quantitative data were gathered, being the first composed of pilots comments and Cooper-Harper ratings, and the latter of information derived from the parametric methodology proposed by the QLC, more specifically, the number of envelopes crossed.

IV. Results and Discussion

In this Section, respectively by the observation of Cooper-Harper ratings and number of envelopes crossed according to the QLC, qualitative and quantitative results are given and a correlation analysis between them is conceived. In a reduced scope of several possible statistical analysis, at this point, data are observed from the perspective of an overall preliminary result, therefore, it is *not* the intend to individually look at each pilot and/or seek for differences between them. Numerical significance is provided to the correlation analysis by the measurement, via a number called *correlation coefficient*, of the strength of the relationship between the variables *Cooper-Harper ratings* and *number of envelopes crossed*. It is relevant to mention that these are understood as *ordinal variables*, since both the rating scale and the QLC establish an order and categorise their outputs, which unfolds in the fact that statistical mean and standard deviation may not be appropriate to represent central tendency and variability of the ordinal variables, therefore, median and interquartile range are used, instead [25]. Prior to the presentation of the results in terms of HQRs and QLC outputs, a discussion is posed the pilots' general perception regarding the equipment utilised and the assembly of the simulation experiments.

A. Pilots' general observations about the experiments

Prior to the realisation of the experiments themselves, a total of five pilots, with different backgrounds, were invited to take part in the flight simulator set-up process and also to fine-tune the simulation procedure. Pilots helped to place the sidestick, rudder pedals and engine controls in an ergonomic position within the cockpit and also to adjust their sensitivity curves, moreover, the simulation scenes to be displayed in the screens, including the identification of the information to be exhibited in the instrument panel, were also adjusted. Aircraft loading condition was also fine-tuned, being the *FlightGear* vehicle model behaviour verified to be in accordance with pilots' expectations regarding different combinations of weight and CG position, fact accompanied by the motion of the simulator, which was addressed after adjustments in the parameters of the platform PID motion controller. Still before the experiments, pilots also shared their knowledge and experience to modify and propose tasks to be performed in the simulations, as well as to define the performance criteria to be used in each manoeuvre, that both in consideration of a fully operational aircraft and simulation scenarios.

In a general opinion, following the realisation of the experiments themselves, in which concerns the adopted simulation procedures, the aviators recognised similarities with their daily practices, mainly due to the existence of a preflight briefing explanation section and the possibility of familiarising themselves with the simulator equipment, the test scenarios and also with the tasks. Moreover, the explanation of the Cooper-Harper scale was also positively seen by the pilots and some attested congruence with their background knowledge, despite the recognition that sometimes the scale is used in different ways in their routine jobs.

Regarding the simulation scenarios, the aviators usually mentioned during the preflight briefing the existence of redundancy to mitigate some of the vehicle impairments addressed by the scenarios, especially in terms of loss of elevator effectiveness (Scenario 02) and unresponsive engines (Scenario 03), moreover, some pilots pointed that the modifications on the control curves performed to address exacerbated inputs (Fig. 5) sometimes led to a vehicle response incompatible to that expected for an aircraft of the size and inertia of a Boeing 777. Despite these observations, all pilots attested realism to the simulation scenarios and assimilated the combinations of precursors, some even elongating the preflight briefing section to share their experiences in flight test campaigns regarding similar conditions.

Concerning the conceived tasks, the evaluation pilots attested that they would expect to perform the assigned manoeuvres in a routine operating basis, but these are usually less demanding in normal procedures than the levels of performance proposed in each simulation exercise. Furthermore, pilots mentioned to be indeed immersed in the simulation environment during the realisation of the tests and felt comfortable with the simulator aircraft model response, as well as the platform motion. It is noteworthy, however, that some pilots eventually reached the platform saturation limits (Table 1) during the execution of a certain task and that the simulator pitch and roll maximum angles are smaller than those understood as usual attitude borderlines and incorporated by QLC UA envelpe.

B. Cooper-Harper Ratings

The handling qualities results for the tasks performed in the piloted simulations are summarised in Fig. 6 in the form of a box and whisker plot of the numerical ratings assigned by six different aviators. The low number of HQRs outliers (approximately, 6%) attests the coherence of the methodology employed in the simulations in the sense of reducing pilot variability, an impossible to remove aspect of every handling quality experiment due to differences in piloting strategies and subjectivity of the *compensation* wording [19].

In an overview, Fig. 6 shows that handling qualities in Scenario 01 are usually rated Level 1 (HQRs *1* to *3*), since an OEI situation is not supposed to push flying qualities to poor conditions, furthermore, given that pilots are usually trained for an OEI flight, there is no crew inappropriate response precursor among those of Scenario 01. The second combination of contributing factors (Scenario 02), in turn, is typically associated with Level 2 ratings (HQRs *4* to *6*), as well as the only task performed in Scenario 03; in common, these sets share the exacerbated crew inputs LOC-I precursor. Scenario 04, on the other hand, is more diverse as Level 2, Level 3 (HQRs *7* to *9*) and even Level 4 (HQR *10*) ratings are encountered, meaning that handling qualities are indeed deeply connected to the manoeuvre being performed [14], in fact, it is attested by the fact that tasks addressing the capability of changing the aircraft longitudinal

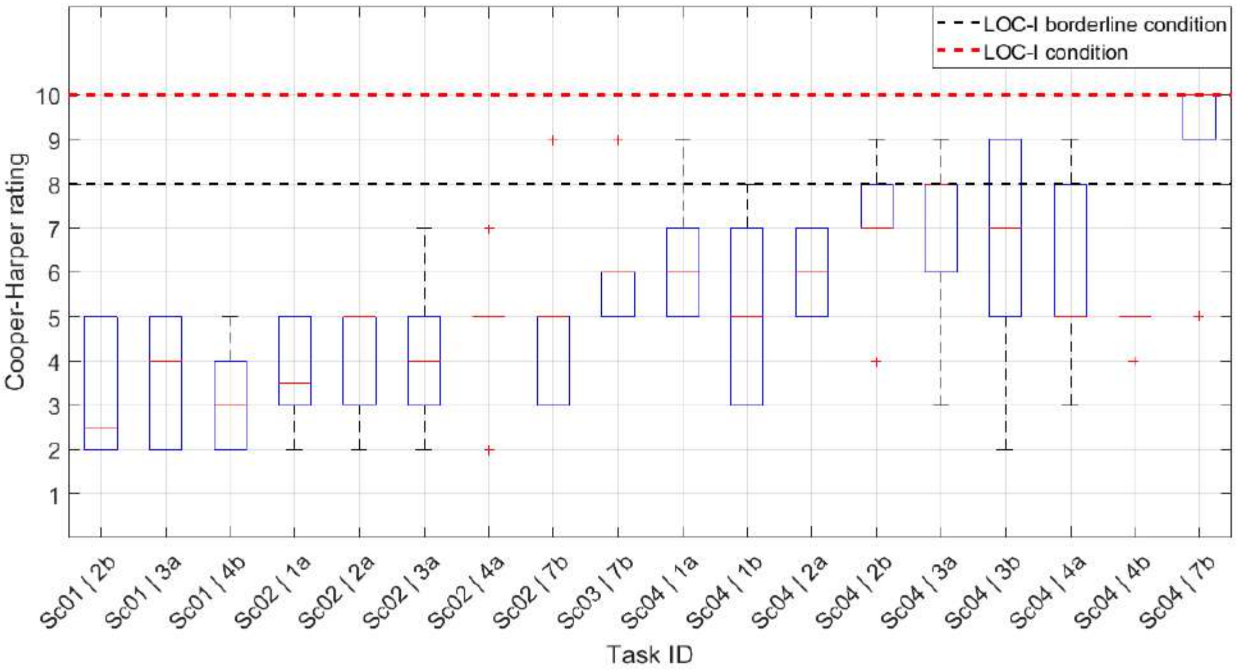


Fig. 6 Cooper-Harper ratings.

attitude (tasks *3a* and *3b*) are worse rated than those associated with the maintenance of a certain pitch condition (tasks *1a* and *1b*), meanwhile the opposite occurs for the lateral-axis exercises[**], which is probably associated with differences in power requirements posed by the manoeuvres and the necessity of compensating yaw tendencies (bearing in mind the degraded condition of the vehicle due to icing). Finally, the fourth set of precursors is the only combination of precursors addressed in this research yielding to produce a LOC-I condition (HQR *10*), or imminence of such (HQRs *8* and *9*), in pilots' perspective.

From the observation of manoeuvres performed in more than one set of precursors, Fig. 6 also points that, in a general perspective, Scenarios 01 to 04 are arranged in an ascending order of pilot degree of compensation. In fact, it is important to remember that total pilot workload is comprised of the workload due to the task itself plus the compensation, moreover, the decision tree structure of the Cooper-Harper scale (thoroughly followed in the experiments) prioritises the assessment of pilot workload, instead of performance criteria, thus, a same task yielding in HQRs of different Levels according to the simulation scenario lead to the conclusion that the degree of compensation is indeed the reason for the difference. Taking as example task *7b*, a summary of pilots debriefing comments confirm it:

- **Scenario 02:** pitch attitude, as well as bank angle control, is fair, even though the aircraft in pitch be sluggish and, in roll, tending towards something abrupt. Displacement of the sidesitck, both in the longitudinal and lateral axes, is moderate, the same valid for its frequency of use. Rudder pedals are not used by the majority of pilots, and airspeed control is somewhere in between easy and fair. No PIO tendency nor special control technique required. Physical and mental workload varying from small to moderate;
- **Scenario 03:** pitch attitude and bank angle control are considered to be from fair to difficult and aircraft response, both in pitch and roll, is sluggish, but also abrupt. Sidesitck displacements tend to be small, but its frequency of use is high. Rudder pedals are not used and the airspeed control lies in between fair and difficult. Special control techniques are employed aiming to avoid a possible longitudinal PIO. Localizer tracking did not pose problems, but following the glideslope was difficult. Physical and mental workload stay between moderate to high;

[**] That is, tasks related to the evaluation of the capability of changing the aircraft lateral attitude (tasks *4a* and *4b*) typically have better HQRs than those associated with the maintenance of a bank angle condition (tasks *2a* and *2b*)

- **Scenario 04:** both pitch attitude and bank angle control are difficult for most pilots, being the longitudinal and lateral aircraft responses sluggish and unpredictable. Control column displacements are wide and used at a high frequency. Pedals are eventually used and the airspeed control is difficult (some pilots even said it to be "impossible"). Special techniques are used, as some pilots tried to compensate the delay in roll response by using pedal inputs, meanwhile others used throttle levers asymmetry, moreover some still combined high gain and small displacement actuation on the sidestick to avoid a severe longitudinal PIO tendency. Physical and mental workload are high.

Finally, the encounter of HQRs corresponding to controllability thresholds or even LOC-I conditions in the simulation experiments, states that the preliminary LOC-I test scenarios proposed in [23] and adapted for the purposes of this research, are effective in their objective to be used for testing LOC-I conditions (that in pilots' perspective). Beyond it, bearing in mind that the tasks performed in the simulations are driven by real-world demands, the finding unfolds in the fact that the presence in an operational environment of the precursors addressed by the scenarios may result in accidents, as past events sadly attest.

C. Number of envelopes crossed – QLC

Following the parametric methodology suggested by the QLC, data of the various tasks performed by the six different pilots were plotted accordingly (additional data required to normalise variables used by the QLC are presented in Appendix B) to determined the number of envelopes crossed, being the results presented in Fig. 7. Again, few outliers occur (approximately 2%), showing that the methodology consistency also reverberates in the quantitative approach, moreover, on average, no envelopes are extrapolated in Scenarios 01 and 03, and for the sets 02 and 04, one envelope is usually crossed, which, in turn, does not represent a category shift for the quantitative method. Figure 7 still shows that, despite some pilots have eventually crossed two or even three envelopes, only in task *7b* from Scenario 04 there is convergence to a LOC-I borderline condition (comparatively, this is the only manoeuvre consistently rated Level 4 in pilots' opinion according to the Cooper-Harper scale), attesting that, in the perspective of the QLC, the test scenarios rarely resulted in LOC-I or even *near*-LOC-I conditions, that both for an overall result between different pilots and even individually looking at each aviator. Furthermore, the ascending pilot degree of compensation from Scenario 01 to 04 does not have correspondence on the QLC, as little to no difference is seen between the number of envelopes excursed across the simulation scenarios.

In order to understand the impact of aircraft deficiencies on the parametric methodology, a more detailed investigation shows that the only envelope excursed in Scenario 01 is the DRC and it occurred for the lateral tasks – namely, *2b* and *4b* –; in Scenario 02, DRC crossing occurs for the exercises majorly involving the lateral axis – *2a* and *4a*. The landing approach (*7b*) in Scenario 03 resulted in the excursion of the DPC and AA envelopes, although not resulting in a stall; finally, in Scenario 04, the dynamics envelopes (DPC and DRC) again represent the most common crossings, and the correspondence between longitudinal manoeuvre and DPC extrapolation, as well as lateral-axis exercise and DRC excursion, is valid, however, there is no significant difference between tasks evaluating the maintenance of a certain attitude (tasks *1a*, *1b*, *2a* and *2b*) and the capability of changing the flight condition (tasks *3a*, *3b*, *4a* and *4b*).

Although not representing a difference to the quantitative criteria, the identification of the envelopes typically crossed shows that AA, UA and SI excursions are rare and it is possibly associated to the conception of the tasks, process partially driven by the restrictions imposed to guarantee the proper use of the Cooper-Harper rating scale. In further details, to the qualitative assessment, pilots must be aware of the parameters and tolerances defining the performance criteria [16], therefore, tasks were designed in a way that pilots could monitor the evolution of these parameters during the execution of the manoeuvres and then actuate on the controls to counteract any deviations from the assigned exercise. Due to such necessity, the variables under consideration to conceive *performance* (as for the scale) are essentially those provided on the aircraft PFD – i.e., airspeed, altitude, pitch angle and bank angle – and some of these form the basis of the UA and SI envelopes. In the perspective of these mentioned envelopes, the assigned tasks typically consists of taking the aircraft close to their borderlines, but the achievement of the wider level of performance (*adequate* condition) guarantees no excursion, given that it would be an inconsistency with a safe flight definition to consider as adequate the excursion of usual pitch and bank attitudes, for example. Hence, by monitoring the parameters under analysis for a given task and attaining to the adequate level of performance, pilots are automatically avoiding the extrapolation of

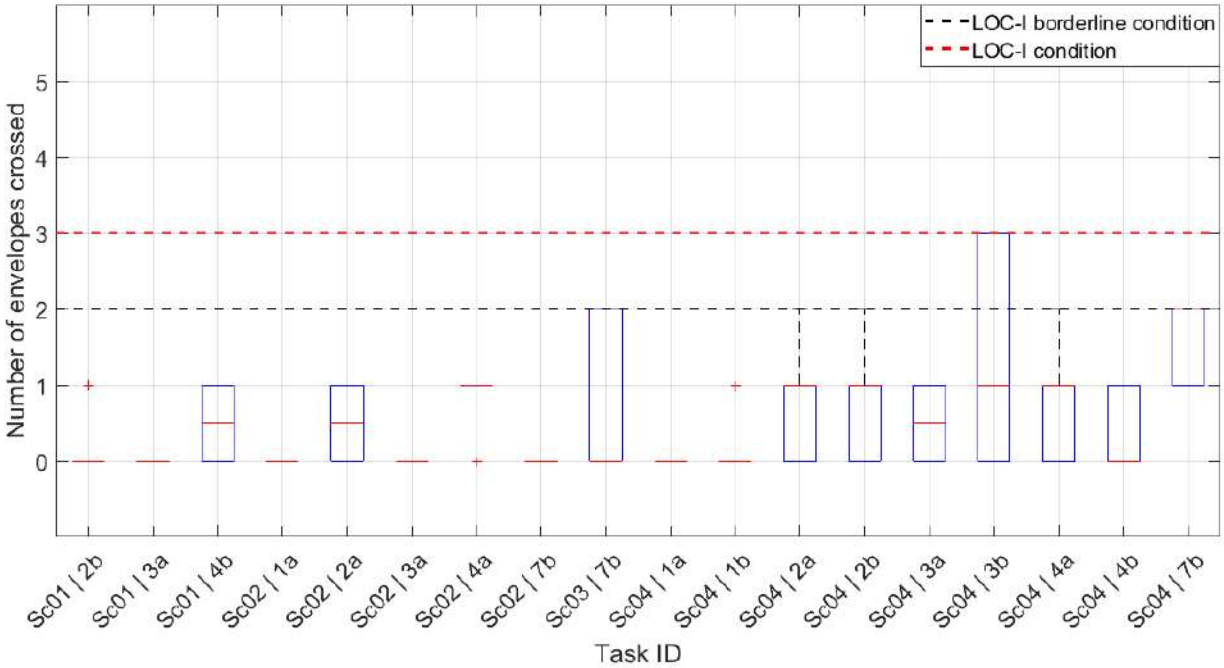


Fig. 7 Number of envelopes crossed according to the QLC.

two[††] of the five QLC envelopes.

D. Correlation

Following the results obtained for the Cooper-Harper ratings and number of envelopes crossed, the correlation between them is investigated. In fact, more precisely, the analysis is conducted for the classification categories resulting from the qualitative ratings and the amount of envelopes extrapolated, i.e., normal situations correspond to HQRs from *1* to *7*, controllability threats (or *near*-LOC-I) to ratings *8* and *9* and, finally, HQR *10* states an indeed LOC-I; respectively, on the QLC perspective, normality is represented by zero or one envelope crossed, *near*-LOC-I by two crossings and, lastly, three or more excursions are classified as a LOC-I.

As a consequence of the category variables being ordinal, as well as the lack of indication of how such a relationship, if any, takes place, and uncertainties surrounding the normality of data distribution, the *Spearman's rank correlation coefficient* – r_S coefficient – is selected as the correlation parameter. For instance, the r_S coefficient assesses the existence (or not) of a monotonic relationship between the mentioned variables, details on how the exchange takes place (linear, exponential etc) are not provided [26]. Equation 5 refers to the calculation of the Spearman's correlation coefficient for data with considerable number of rank ties, being N the total sample size, $R(x)$ and $R(y)$ the rankings for x and y data and $\overline{R(x)}$ and $\overline{R(y)}$ the mean ranks for x and y [27]; x and y represent, respectively, the categories resulting from the assigned HQRs and number of envelopes crossed.

$$\mathbf{r_S} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\left(R(x_i)-\overline{R(x)}\right)\left(R(y_i)-\overline{R(y)}\right)\right)}{\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\left(R(x_i)-\overline{R(x)}\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left(R(y_i)-\overline{R(y)}\right)^2\right)}} \tag{5}$$

[††] In terms of the SI envelope, velocity extrapolation, at least

Table 6 presents the correlation coefficient calculated in consideration of all simulation scenarios and data gathered with the six different pilots, thus the sample size adds up to $N = 108$; moreover, assuming a null-hypothesis that there is no monotonic correlation (and, consequently, an alternative hypothesis that there is monotonic relationship), a p_{value} is calculated and also provided in Table 6.

Table 6 Spearman's correlation coefficient, p-value and sample size for the relationship between categories resulting from *Cooper-Harper ratings* and *number of envelopes crossed*.

	All Scenarios
r_S **coefficient**	+0.4682
p-value	3.23e-07
N (sample size)	108

The obtained value for the correlation coefficient ($r_S = 0.4682$) indicates that a positive correlation between the investigated variables exists and that it is statistically significant (p_{value} = 3.23e-07, assuming a significance level of 0.01 – two-tailed); no information is provided about the type of relationship between the variables. Even though with evidences to refuse the null-hypothesis, the relationship is considered from weak to moderate due to the magnitude of the correlation parameter. It is relevant to observe that the calculation was based on all simulations performed and, among these, all categories resulting from the Cooper-Harper scale and the QLC ('normal', '*near*-LOC-I' and '*LOC-I*') are present, but the events of interest (those indeed impacting on controllability) are significantly less frequent in both approaches, therefore, the correlation is essentially driven by events in the normality range. Figure 8 shows the frequency of events in each of the mentioned categories resulting from the application of both methods; the dominance of manoeuvres classified as 'normal' is clear, independently of the approach.

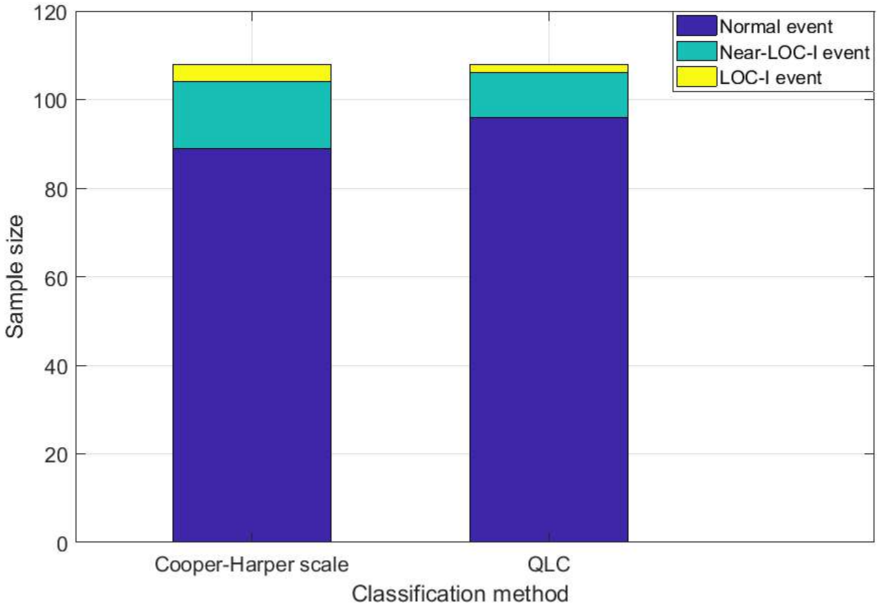


Fig. 8 Frequency of events according to their controllability categories resulting from the application of the classification methods.

The small number of *near*-LOC-I and LOC-I events verified in both approaches, as well as the apparent lack of correspondence on the QLC for the growing level of compensation seen in pilots' perceptive, may result from inherent characteristics of the selected qualitative and quantitative methods. In the Cooper-Harper rating scale, controllability is assessed *within the task*, thus, different manoeuvres performed in the same simulation scenarios, and possibly even more representative of real-life conditions, could result in more occurrences rated *10*, furthermore, despite many efforts to provide more objectivity to the compensation wording used in the scale, the term *maximum tolerable workload*, decisive for distinctions between normal events (HQRs *1* to *7*) and *near*-LOC-I conditions (HQRs *8* and *9*) is (and will ever be in any Cooper-Harper assessment) subjective and a source of intra and interpilot variabilities (for a same pilot, fatigue, for example, may produce different understandings of the *tolerable* workload level).

Regarding the QLC, when conceiving the test scenarios, [23] points that *vehicle upsets*, like unusual attitudes and stall, are common factors within the chain of events leading to a LOC-I condition, however, these are rarely seen in the plots resulting from simulations. The conception of the tasks made difficult the occurrence of critical upsets in the experiments, essentially because the manoeuvres were based on parameters directly impacting on the QLC, like those conceiving the UA and SI envelopes, although, it must be pointed that the test scenarios simulate a fault vehicle (e.g., there is a significant loss of elevator effectiveness in Scenario 02), but these circumstances are not accounted for by the borderlines of the envelopes, since they are conceived for a fully operational aircraft and applied as if the vehicle presented no failures. [28] discusses about the necessity of adapting QLC borderlines to permissible flight envelopes according to the aircraft condition and, as a consequence of incorporating these changes, it is predicted that a more thorough correlation between the qualitative and quantitative approaches may be conceived, as *near*-LOC-I and LOC-I events become statistically more representative. Finally, it is remarkable that none of the QLC envelopes suggest a connection with pilot degree of compensation; such variable has a strong importance for the Cooper-Harper scale and actually drives the classification procedure, therefore, metrics as pilot control deflection rates, representative of mental pilot workload [24], should be investigated as a possible addition to the criteria.

V. Concluding Remarks

In conclusion, it is possible to attest that the correlation between human pilot perspective – by means of the Cooper-Harper rating scale – and aircraft behaviour – using the Quantitative Loss of Control Criteria – has been conceived following the application of a comprehensive methodology for the study of LOC-I events using pilot-in-the-loop simulations. After performing experiments with a total of six different pilots and nine distinguished manoeuvres, both methods yielded to produce three different possible conditions: (i) controllability was not in question; (ii) a LOC-I borderline and; (iii) a LOC-I, however, only in pilots' perspective these three categories are consistently present.

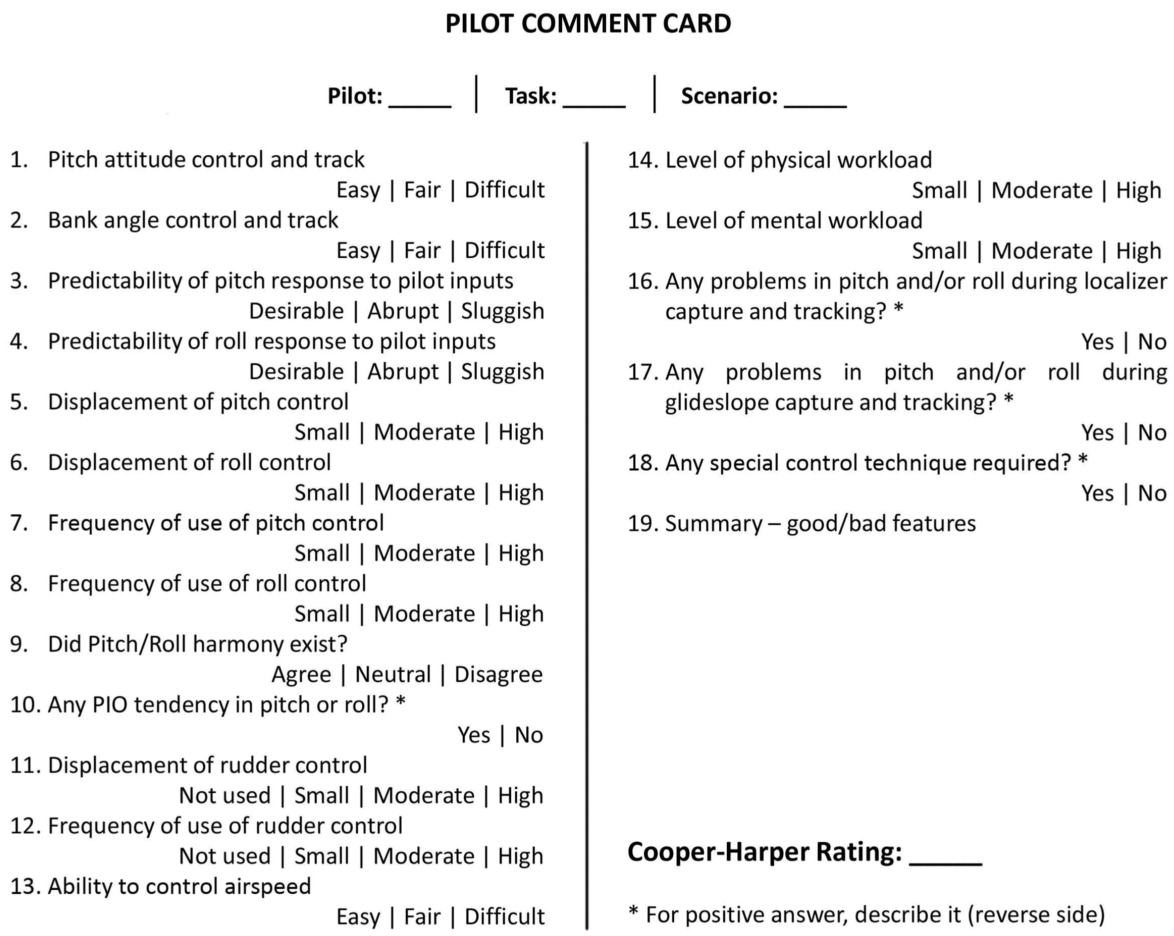
Still, conditions within the normality range were the most frequent in both approaches, pointing that the existence of a positive weak to moderate correlation between the methods is not much representative due to the lack of data within the other categories and also attesting that the test scenarios 03 and 04 were more effective to address the LOC-I issue than sets 01 and 02. To enhance the chances of effectively reaching and observing control problems, adaptations to the set of tasks and controllability assessment were proposed, essentially laying on (i) the suggestion of conceiving manoeuvres to be performed by the pilots based on parameters *not* directly linked to a QLC envelope and (ii) the necessity of adapting QLC borderlines according to the aircraft condition, instead of using parameters relative to a fully operational vehicle.

Finally, it is important to bear in mind that engineering intervention strategies are usually based on quantitative metrics and strategies to mitigate LOC-I accidents inevitably go through interfering in the human-machine interaction, thus, to the future, the incorporation by the QLC of a variable relating to pilot compensation should be investigated, since the factor may well represent human being limitations, leveraging the findings to be used for the development of defences capable of considering, together, two of the major participants in LOC-I events and effectively mitigate the accident.

Appendix

A. Pilot comment card

Figure 9 presents the comment card used during simulations. The card is a sort of "open-questionnaire", since the questions were designed in a way to combine different qualities in a single questioning, opening the possibility to assess pilots' objections and their causes. The inquiries fostered pilots to consider factors of multiple orders and then conceive a more complete analysis, moreover, it also allowed the engineer to observe whether the objectives of the experiment were actually realised, if the terms used in the Cooper-Harper scale were accurately interpreted by the pilots and if the proposed tasks were well-designed and clear, therefore, pilots' answers tended to be open and long, instead of simple "YES" or "NO". This way, the options presented after each question in Fig. 9 merely worked as guides and *not* as truly accepted answers.



PILOT COMMENT CARD

Pilot: _____ | **Task: _____** | **Scenario: _____**

1. Pitch attitude control and track
   Easy | Fair | Difficult
2. Bank angle control and track
   Easy | Fair | Difficult
3. Predictability of pitch response to pilot inputs
   Desirable | Abrupt | Sluggish
4. Predictability of roll response to pilot inputs
   Desirable | Abrupt | Sluggish
5. Displacement of pitch control
   Small | Moderate | High
6. Displacement of roll control
   Small | Moderate | High
7. Frequency of use of pitch control
   Small | Moderate | High
8. Frequency of use of roll control
   Small | Moderate | High
9. Did Pitch/Roll harmony exist?
   Agree | Neutral | Disagree
10. Any PIO tendency in pitch or roll? *
   Yes | No
11. Displacement of rudder control
   Not used | Small | Moderate | High
12. Frequency of use of rudder control
   Not used | Small | Moderate | High
13. Ability to control airspeed
   Easy | Fair | Difficult
14. Level of physical workload
   Small | Moderate | High
15. Level of mental workload
   Small | Moderate | High
16. Any problems in pitch and/or roll during localizer capture and tracking? *
   Yes | No
17. Any problems in pitch and/or roll during glideslope capture and tracking? *
   Yes | No
18. Any special control technique required? *
   Yes | No
19. Summary – good/bad features

Cooper-Harper Rating: _____

* For positive answer, describe it (reverse side)

Fig. 9 Pilot comment card used during simulation experiments.

B. Additional data required by the QLC

As detailed in Section II.B, the QLC requires the normalisation of the airspeed and the angles of attack and sideslip, however, this information is not promptly available in manuals and are dependent on specific aircraft configuration. Using similar procedures as those employed in real flight test campaigns, tests were conducted in EESC-USP's simulator to determine the normalisation quantities; in summary, considering the only two aircraft configurations used in the experiments, they are presented in Table 7.

The angle of attack for stall warning activation, α_{SW}, was approximated as the angle corresponding to the beginning of the prestall region, point marked by aerodynamics nonlinearities as well as a nonzero roll rate tendency [29]; β_{MDXW}, on the other hand, was determined following approaches performed in the maximum manufacturer's demonstrated takeoff/landing crosswind component, 45 kts and, finally; the maximum operating equivalent airspeed for flaps up and flaps down configurations (V_{FE} and V_{MO}, respectively), as well as the stall warning equivalent airspeed in 1-g flight, were determined according to airspeed tape indications provided on the PFD.

Table 7 Parameters used to normalise QLC quantities.

Aircraft Configuration	Parameter	Value
Flaps retracted LDG up	α_{SW}	12.8
	β_{MDXW}	7.4
	V_{MO}	330
	V_{SW}	200
Flaps full LDG down	α_{SW}	12.8
	β_{MDXW}	7.4
	V_{FE}	170
	V_{SW}	120

Acknowledgments

We sincerely thank the pilots for participating in the research, who very willingly shared their knowledge and devoted their attention to contributing to the study. This research was partially supported by the National Council for Scientific and Technological Development (CNPq).

References

[1] International Air Transport Association, "Loss of Control In-Flight Accident Analysis Report 2010-2014," Tech. rep., International Air Transport Association, Montreal, 2015. URL `https://www.iata.org/whatwedo/safety/Documents/LOC-I-1st-Ed-2015.pdf`.

[2] Belcastro, C., Newman, R., Crider, D., Klyde, D., Foster, J., and Groff, L., "Aircraft Loss of Control: Problem Analysis for the Development and Validation of Technology Solutions," *AIAA Guidance, Navigation, and Control Conference*, American Institute of Aeronautics and Astronautics, San Diego, 2016, pp. 1–48. https://doi.org/10.2514/6.2016-0092, URL http://arc.aiaa.org/doi/10.2514/6.2016-0092.

[3] International Air Transport Association, "Safety Report 2016," Tech. rep., International Air Transport Association, Montreal, 2017. URL http://www.iata.org/publications/Pages/safety-report.aspx.

[4] Airbus, "A Statistical Analysis of Commercial Aviation Accidents 1958-2016," Tech. rep., Airbus Societas Europaea, Toulouse, 2017. URL https://skybrary.aero/bookshelf/books/4009.pdf.

[5] International Air Transport Association, "Safety Report 2013," Tech. rep., International Air Transport Association, Montreal, 2014.

[6] International Air Transport Association, "Safety Report 2014," Tech. Rep. April, International Air Transport Association, Montreal, 2015.

[7] International Air Transport Association, "Safety Report 2015," Tech. Rep. April, International Air Transport Association, Montreal, 2016. URL http://www.iata.org/publications/Documents/iata-safety-report-2015.pdf.

[8] Lambregts, A., Nesemeier, G., Wilborn, J., and Newman, R., "Airplane Upsets: Old Problem, New Issues," *AIAA Modeling and Simulation Technologies Conference and Exhibit*, American Institute of Aeronautics and Astronautics, Honolulu, 2008, pp. 1–10. https://doi.org/10.2514/6.2008-6867, URL http://arc.aiaa.org/doi/10.2514/6.2008-6867.

[9] National Transportation Safety Board, "NTSB 2017-2018 Most Wanted List of Transportation Safety Improvements," Tech. rep., National Transportation Safety Board, Washington DC, 2016. URL https://www.ntsb.gov/safety/mwl/Documents/2017-18/MWL-Brochure2017-18.pdf.

[10] Boeing Commercial Airplanes, "Statistical Summary of Commercial Jet Airplane Accidents - Worldwide Operations | 1959 – 2016," Tech. rep., Boeing Commercial Airplanes, Seattle, 2017. URL www.boeing.com/news/techissues/pdf/statsum.pdf.

[11] European Aviation Safety Agency, "European Plan for Aviation Safety 2016-2020," Tech. rep., European Aviation Safety Agency, Cologne, 2016. URL https://www.easa.europa.eu/sites/default/files/dfu/EPAS2016-2020FINAL.PDF.

[12] International Civil Aviation Organization, "Doc 10004 Global Aviation Safety Plan 2017-2019," Tech. rep., International Civil Aviation Organization, Montreal, 2016. URL https://www.icao.int/safety/Pages/GASP.aspx.

[13] Bromfield, M., and Landry, S., "Loss of Control In Flight (LOC-I) - time to re-define?" *AIAA Aviation 2019 Forum*, American Institute of Aeronautics and Astronautics, Dallas, 2019. https://doi.org/10.2514/6.2019-3612, URL https://arc.aiaa.org/doi/abs/10.2514/6.2019-3612.

[14] Cooper, G., and Harper Jr., R., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," , 1969.

[15] Wilborn, J., and Foster, J., "Defining Commercial Transport Loss-of-Control: A Quantitative Approach," *AIAA Atmospheric Flight Mechanics Conference and Exhibit*, American Institute of Aeronautics and Astronautics, Providence, USA, 2004, pp. 1–11. https://doi.org/10.2514/6.2004-4811, URL http://arc.aiaa.org/doi/10.2514/6.2004-4811.

[16] Mitchell, D., “Fifty years of the cooper-harper scale,” *AIAA Scitech 2019 Forum*, American Institute of Aeronautics and Astronautics, San Diego, 2019, pp. 1–18. https://doi.org/10.2514/6.2019-0563, URL https://arc.aiaa.org/doi/abs/10.2514/6.2019-0563.

[17] European Aviation Safety Agency, “Loss of control prevention and recovery training,” , 2017. URL https://www.easa.europa.eu/sites/default/files/dfu/OpinionNo06-2017.pdf.

[18] International Civil Aviation Organization, “Doc 10011: Manual on Aeroplane Upset Prevention and Recovery Training,” , 2014. URL https://www.icao.int/Meetings/LOCI/Documents/10011{_}draft{_}en.pdf.

[19] Wilson, D. J., and Riley, D. R., “Cooper-harper pilot rating variability,” *16th Atmospheric Flight Mechanics Conference, 1989*, American Institute of Aeronautics and Astronautics, Boston, 1989, pp. 96–105. https://doi.org/10.2514/6.1989-3358.

[20] Lemes, R. C., Souza, M. M., Belo, E. M., and Bidinotto, J. H., “Latency on a Stewart platform using washout filter,” *Aeronautical Journal*, Vol. 122, No. 1252, 2018, pp. 1003–1019. https://doi.org/10.1017/aer.2018.35.

[21] Hollnagel, E., Woods, D., and Leveson, N., *Resilience Engineering: Concepts and Precepts*, Ashgate Publishing Limited, Hampshire, 2006. https://doi.org/10.1136/qshc.2006.018390.

[22] Reason, J., *Human Error*, Cambridge University Press, New York, 1990.

[23] Belcastro, C., “Validation of Safety-Critical Systems for Aircraft Loss-of-Control Prevention and Recovery,” *AIAA Guidance, Navigation, and Control Conference*, American Institute of Aeronautics and Astronautics, 2012, pp. 1–30. https://doi.org/10.2514/6.2012-4987, URL http://dx.doi.org/10.2514/6.2012-4987.

[24] Lombaerts, T., Smaili, M., Stroosma, O., Chu, Q., Mulder, J., and Joosten, D., “Piloted Simulator Evaluation Results of New Fault-Tolerant Flight Control Algorithm,” *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 6, 2009, pp. 1747–1765. https://doi.org/10.2514/1.44280, URL https://doi.org/10.2514/1.44280.

[25] Hodgkinson, J., *Aircraft Handling Qualities*, 1$^{\text{st}}$ ed., Blackwell Science Inc., Oxford, UK, 1999.

[26] Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K., *Probability and Statistics for Engineers and Scientists*, ninth ed., Pearson Education, Inc., Boston, MA, 2011.

[27] Cleff, T., *Exploratory Data Analysis in Business and Economics*, Springer, Pforzheim, GE, 2014. https://doi.org/10.1007/978-3-319-01517-0.

[28] Pfifer, H., Venkataraman, R., and Seiler, P., “Quantifying loss-of-control envelopes via robust tracking analysis,” *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 4, 2017, pp. 1042–1050. https://doi.org/10.2514/1.G001748, URL https://doi.org/10.2514/1.G001748.

[29] Richards, N. D., Gandhi, N., Bateman, A. J., Klyde, D. H., and Lampton, A. K., “Vehicle upset detection and recovery for onboard guidance and control,” *Journal of Guidance, Control, and Dynamics*, Vol. 40, No. 4, 2017, pp. 920–933. https://doi.org/10.2514/1.G001738, URL https://doi.org/10.2514/1.G001738.