

An Evaluation of Several Stall Models for Commercial Transport Training

Jeffery A. Schroeder¹

Federal Aviation Administration, Moffett Field, California, 94035

Judith Bürki-Cohen²

John A. Volpe National Transportation Systems Center, Cambridge, Massachusetts, 02142

David A. Shikany³

Boeing Commercial Airplanes, Seattle, Washington, 98124

David R. Gingras⁴

Bihrlle Applied Research, Hampton, Virginia, 23666

Paul Desrochers⁵

Test Pilot, inc, Weatherford, Texas, 76088

Abstract

Air carriers will use simulators to train pilots to recover from fully developed stalls. Flight simulator models will have to portray an airplane's dynamics satisfactorily to achieve the associated training objectives. That imposes new simulator requirements. This study evaluated several full stall simulator models to meet those requirements. Three new stall models were tested in a Boeing 737-800 simulator. One model used the conventional approach by matching flight test data to within a tolerance. Another model did not rely on flight test data, but instead combined computational aerodynamics, scaled wind tunnel data, and expert opinion from a test pilot who had stalled the actual aircraft. The third model added a roll asymmetry to the unmodified simulator model as a simple way to possibly meet the training objectives. The test had two phases. In the first phase, test pilots who had stalled a 737 airplane evaluated the models by performing typical flight test stall maneuvers in the simulator. The second phase used airline pilots type-rated in the 737 but who had not stalled a 737 airplane. The airline pilots were placed in groups, and each group trained with one of the models. Each airline pilot was then checked on the model developed from flight data, which represented the truth model. The second phase also included a surprise stall scenario with each airline pilot having to recover from a stall using the model they would train with. The results revealed wide ranges in the subjective evaluations of the test pilots, as well as in the objective performance of the airline pilots across the models. However, many of the averages did not show significant differences. All airline pilots agreed or strongly agreed that they were surprised by the surprise stall scenario. In that scenario, less than one quarter of the airline pilots strictly followed the proper stall recovery procedure on which they had been briefed. Less than half maintained a nose-down input until the stall warning stopped. For situations when developing a stall model based on flight data is not practical, the alternative approach of developing a model based on computational aerodynamics, wind tunnel data, and subject expert opinion appears feasible.

Introduction

By early 2019, pilots that fly for air carriers in the United States will receive training to recover from fully developed stalls.¹ Except in rare cases, the training will occur in flight simulators. Typically, an airplane simulator model is built to match the actual airplane response within certain tolerances.^{2,3} However, fully developed stall data may be difficult to obtain for all commercial aircraft. As such, a committee of subject matter experts proposed that stall recovery training objectives could be met with a model that was not developed from flight data.⁴⁻⁷ Their belief

¹ Chief Scientific and Technical Advisor for Flight Simulation Systems, Mail Stop 243-5, Associate Fellow.

² Principal Technical Advisor, USDOT-RITA-Volpe Center-RVT-81, 55 Broadway, Senior Member

³ Associate Technical Fellow, Flight Simulation Aerodynamics Stability and Control.

⁴ Vice President, 81 Research Drive. Senior Member.

⁵ President, PO Box 277.

was that a satisfactory model could be built from analysis, wind tunnel data, and input from a pilot who has stalled the aircraft. This reasoning is hypothetical; data do not exist to confirm it.

Creating models for fully developed stall behavior made great strides from the combined efforts of NASA and Boeing.⁸ Stall models have recently been created for a hypothetical airplane.⁹⁻¹¹ and for estimating a 747's stall characteristics¹² to investigate cueing effects in stalls and other upset recoveries. The potential value of building and using a stall model without reliance on flight data was assessed for U.S. Navy P-8A applications.¹³ What this paper adds to extant literature is a comparison of several stall modeling approaches for the purpose of satisfying civil transport training objectives. These comparisons employed a large group of test pilots who had stalled the airplane, and it had a large group of airline pilots train with the various models to determine if model-dependent differences might arise.

The paper unfolds as follows. First the experimental set-up is described. This is broken into two phases. One phase was the subjective evaluation by test pilots. The other phase was an objective training evaluation by airline pilots. A surprise stall scenario was also inserted in the airline pilot phase. The set-up is followed by results, which is followed by a discussion. The paper closes with the main findings.

Experimental Set-up

General

This section first presents the general components of the experiment. The experiment had two phases. The first phase was a subjective evaluation of the stall models using test pilots who had stalled a 737 aircraft. The second phase was an objective evaluation using airline pilots with a 737 type rating. Particulars related to each phase follow the general set-up.

Simulator. The study used the FAA's Level D 737-800 simulator in Oklahoma City. It was manufactured by CAE, and the cockpit had an American Airlines configuration. The displays, motion (except the buffet cues), and sound cues were unmodified. The rate of change of the aircraft's elevator feel shift system, which increases the column forces near the stall angle of attack, was increased to better match the aircraft, according to test pilot experience and data from the aircraft manufacturer. The simulator's instructor operating station was modified to show the aircraft's load factor, angle-of-attack, and sideslip.

Modifications to the stall buffet were made by the project's test pilot who had extensive stall experience in the airplane. Sometimes buffet levels are reduced in a simulator to extend component life; however, the desire in this experiment was to match the actual buffet as best as possible without changing the simulator's software structure. The overall buffet gain was increased by a factor of 2.5. The maximum buffet amplitude limit was increased from 1.1 to 3.0. The increased buffet gains that occur in deep stall were invoked 1.2 degrees sooner than in the original simulator model. The flaps 5 onset buffet was reduced from 21.5 degrees to 19.0 degrees. The stall buffet gain was reduced from 0.04 to 0.01 at flaps 5.

Models. The original model, unmodified except for buffet and elevator-feel-shift system cues, was that of a 737-800 without winglets. CAE delivered this model, derived from a Boeing data package, with the simulator. This model was built to satisfy Ref. 2, which focusses on 1-g stall fidelity up to when the stick shaker activates and the quasi-steady responses in the longitudinal axis up to the stall break. Shortcomings have been raised on this model's applicability for full aerodynamic stall training in that it mostly provides one-dimensional stall behavior. That is, the model represents a perfectly rigged symmetric airplane, and its response in a wings-level stall is predominantly in the longitudinal axis, which is consistent with the requirements of the current version of Part 60. Here, this model will be called the "Unmodified" model.

Three new models were created for angles of attack beyond stick shaker activation. Bihrlle Applied Research developed the first new model, which did not rely on flight test data. Instead, it was developed using computational aerodynamics, scaled wind tunnel data from similar aircraft types, and input from a test pilot who had extensive stall experience in many 737 models, including the -800. Incremental stall dynamic effects were added to the "Unmodified" model at angles of attack above stick shaker activation. This process ensured that, above stick shaker, the nonlinear static and dynamic stability & control degradations in the pitch, roll, and yaw axes, which are typical in stalls, are exhibited in a representative form.⁵ Also, unlike the Unmodified model, the Representative model had randomized asymmetry components added in roll and yaw. These components were adjusted using comments by an expert test pilot to represent the full range of aircraft responses experienced during flight test stalls. The model was designed to meet the soon-to-be-published specification by the International Committee for Aviation Training in Extended Envelopes, which is being considered for adoption both by the FAA and the ICAO in revisions

to Refs 2 and 3, respectively. The conjecture is that the added model degradations near stall will allow the objectives of teaching the proper stall recovery technique to be met even if the model was not matched to, or developed from, flight test data. Here, this model will be called the “Representative” model.

Boeing developed the second new model using 737-800 flight test data from hundreds of fully developed stalls, including turning flight stalls not previously used in the original 1990’s flight update of the Unmodified model. At angles-of-attack above stick shaker activation, this model refined the static stability, dynamic stability, and control effectiveness components of the Unmodified model using all available stall flight test data. This process ensured that the new model exhibits characteristics similar to a 737-800 aircraft in a fully developed stall. The model also had a randomized asymmetry component added in roll so that no two stalls are alike in the simulator. This randomness was modified further by Boeing test pilots to match the randomness observed in flight test stalls appropriate to training objectives. As this model was developed using flight test data, it represents the truth model in this experiment. This model will be called the “Specific” model.

The third new model, evaluated only by the test pilots due to time available, had the same Boeing random roll asymmetry added to the “Unmodified” model near the critical angle of attack. The subject matter experts studying the requirements for stall modeling have focused on the fact that simulator models often have one dimensional behavior in a stall, namely in pitch. Adding an asymmetry in roll to the otherwise Unmodified model would require additional pilot workload to control another axis, which is typical of stalls in actual flight. Thus, this model, if acceptable, may offer additional realism at a relatively meager cost. It will be called “Unmodified+roll.” As a caveat, this technique may not be generalizable to all models in the field. Its potential value depends on the extent that an existing model accounts for sufficient control and response degradations near stall. In this instance, the technique was believed to have promise, as the Unmodified model already exhibits some of these degradations.

As noted, randomized asymmetry effects were added to all three of the new models. The roll asymmetry in the Unmodified+roll and the Specific models were the same. The roll and yaw asymmetries in the Representative model were different from the Unmodified+roll and Specific models. These added effects were adjusted by separate sets of experienced test pilots and deliberately not harmonized to assess the possible effects of the differences.

Tasks. The experiment had three tasks, all involving a recovery from a fully developed stall. So that all pilots recovered from a full stall, a tone sounded when the critical angle of attack was reached. Pilots were not to recover until that tone sounded. So that all pilots reached a consistent end state in the recovery, a tone sounded again when they were no longer descending, were at least as fast as their starting speed, and the wings were level. For the airline pilots, to minimize the possibility of negative training, the stick shaker was turned off as they entered the stall (so as not to practice ignoring a stall warning). The shaker was reactivated when the tone first sounded at the critical angle of attack.

Table 1 shows the starting conditions for each task. The winds were calm and the visibility unrestricted. The turbulence was set to 2%. The first task was a turning stall starting at 2,500 ft altitude at maximum landing weight. Pilots rolled to a 30-deg bank angle, established a 1 kt/sec deceleration rate by increasing their pitch attitude until the tone first sounded, and they then recovered until the second tone sounded. The second task was a medium altitude straight-ahead stall at 15,000 ft. Pilots reduced thrust to idle, established a 1 kt/sec deceleration rate until the tone first sounded, and then recovered until the second tone sounded. The third task was a high altitude straight-ahead stall at 39,000 ft at a light weight and aft c.g. Pilots reduced thrust to idle, established a 1 kt/sec deceleration rate until the tone first sounded, and then recovered until the second tone sounded.

Table 1 – Stall training task conditions

Task name	Altitude (ft)	Trim Speed (KCAS)	Flaps/ Gear	Weight (lbs)	c.g %MAC	Thrust
Low	2,500	171	5/Up	144,000	36	~50%
Medium	15,000	213	Up/Up	150,000	36	Idle
High	39,000	201	Up/Up	120,000	36	Idle

These tasks were selected primarily for the reason that the test pilots were likely to have flown them in the actual aircraft, except for the turning stall, which would be performed at a higher altitude for safety. As such, these tasks allowed the test pilots to compare the simulator with their aircraft experience. For simplicity, the airline pilots also flew the first and last tasks, although airlines would likely use an operational context to teach recoveries rather than

use these flight test maneuvers. For instance, airlines may enter the stalls with the autopilot engaged. It is important to stress that it was not the aim of this experiment to serve as an example of an airline training program.

In addition to these training tasks, the airline pilots received a surprise stall scenario. Its intent was to cause an unannounced full aerodynamic stall during a typical line operation. The scenario started with an approach to Reagan National Airport. Pilots flew the RNAV (RNP) RWY 19 arrival with the autopilot engaged but without using a Head-Up Display or autothrottle. The experimenters told the pilots that this was a familiarization task and to expect to hand fly a missed approach when reaching the 480 ft MSL missed approach point. However, they never reached the missed approach point, as a severe tailwind was inserted at 2100 ft MSL. The tailwind was extremely strong, which was necessary, as otherwise pilots would likely recover when noticing a slowing speed into the amber band or when the stick shaker activated. Since the purpose of this study was to evaluate full stall recoveries, a severe external disturbance was needed to reach that condition quickly. The simulator software was changed to ramp in this 74 kt tailwind over 5 seconds. The tailwind did not cause a windshear system warning, as it did not have a vertical shear component.

Table 2 shows the new stall recovery template, introduced in Ref. 14, as developed by the major airframe manufacturers. It emphasizes manually applying nose-down pitch control until the stall warning, which is the stick shaker for this airplane, is eliminated. This principal step of manually reducing angle of attack takes precedence over the subsequent steps of leveling the wings, applying thrust as necessary, and then returning to the desired flightpath. This technique is new. It distinguishes itself from the old recovery technique, which emphasized minimal altitude loss by manually adding thrust.

Table 2. Abbreviated Stall Recovery Template¹

1	Autopilot and auto throttle.....	Disconnect
2	a) Nose down pitch control... ..	Apply until stall warning is eliminated
	b) Nose down pitch trim.....	As Needed
3	Bank.....	Wings Level
4	Thrust	As Needed
5	Speed brakes/Spoilers.....	Retract
6	Return to the desired flight path.	

Both the test pilots and the airline pilots were briefed on this technique prior to flying the simulator. However, the test pilots were told to use the recovery technique they had used in the aircraft, as the purpose of their evaluation was to compare the simulator with the aircraft. The test pilots were asked to keep the Table 2 procedure in mind when assessing each model's appropriateness for learning this technique. The airline pilots were briefed to use this technique, and most pilots stated that their airline was no longer emphasizing minimal loss of altitude.

The airline pilots were being trained to be proficient in recovering from stalls. Proficient was defined using Ref. 15. The airline pilots were rated on their stall recovery technique by the instructor pilot using a stall recovery rating scale shown in the Appendix. The scale is modeled on the Cooper-Harper scale.¹⁶ The descriptions of adequate performance after the scale represent the train-to-proficiency standard in this experiment.

In addition, both the test pilots and the airline pilots were asked to assess several statements using the scale in Table 3. The questions were different for each set of pilots and are given subsequently.

Table 3– Questionnaire Rating Scale

N/A	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
	1	2	3	4	5	6	7

Briefing. Pilots were briefed by the same person prior to their evaluation. The brief described the purpose, the tasks, the stall recovery template, the rating scale, and the questions they were to answer. The test pilots were not told how many different models they would evaluate, but they were told that they would be evaluating the same model for the high and low altitude stalls that were flown in succession. The airline pilots were not told which model they would train with, but they were told that the model would not change during their training runs. The airline pilots were told that they would be checked using the Specific model, which served as the truth model.

Test Pilot Subjective Evaluation

Test Pilots. Nine test pilots participated. Table 4 lists their relevant occupation and their approximate number of stalls in a 737, its equivalent, or other type. One pilot had not stalled a 737 but had E-6 (707) stall experience. Certainly the test pilots have relevant experience. Schedule conflicts prevented participation by current Boeing 737 test pilots, yet many had evaluated the Specific model during its development.

Table 4 – Test pilots

Relevant occupation	Number of 737 stalls
FAA 737NG certification test pilot (retired)	~250
FAA 737NG certification test pilot	~40
FAA test pilot	~12
Boeing 737NG chief test pilot (retired)	>500
Navy P-8A test pilot	>100 P-8A
FAA test pilot	~12
FAA test pilot	None; ~12 on E-6 (707)
Boeing production test pilot (retired)	~200
FAA test pilot	>100

For statistical analysis in this repeated measures design, the null hypothesis was assumed. That is, the assumption is that no matter which of the models the test pilots were flying, it would have no effect on their subjective ratings. Table 5 summarizes the 3x4 experimental matrix for the test pilots using the task names from Table 1. Each task was flown a minimum of three times.

Table 5 – Test pilot experimental conditions

Task	Model			
	Unmodified	Unmod+roll	Representative	Specific
Low	x	x	x	x
Medium	x	x	x	x
High	x	x	x	x

After evaluating each model, each pilot assessed 13 different statements on how the model compared to their flight experience using Table 3 descriptors. The statements are presented later and addressed the handling qualities in pitch, roll, and yaw, as well as key stall characteristics such as buffet and roll-off.

Again, the pilots did not know which model they were flying. The ordering of the low and high altitude stall tasks was balanced. The medium altitude was always flown last (this task was originally going to be flown if time permitted, and time always permitted). The session ended with a debrief.

Airline Pilot Objective Evaluation

Airline Pilots. Forty-five current airline pilots participated. All 45 were type-rated in one of the 737 aircraft series. Of the 45, 18 were captains and 27 were first officers. A total of 28 had previous military experience, and 28 also had aerobatic experience. The 45 pilots flew at nine different carriers: Southwest 19; American 8; Delta 6; United 4; Alaska 2; Air India 2; USAir 1; Sky King 1; and Sun Country 1. The 45 pilots were divided into three groups of 15, and they were divided as evenly as possible to balance backgrounds and experience. Each group was then assigned to train with one of the three models: Unmodified, Representative, or Specific.

This airline pilot part of the experiment had several hypotheses: (1) the groups that trained with either the Representative or Specific model would perform better in the check event than the group that trained with the Unmodified model; (2) the measures for the Representative and Specific models would be the same within operationally relevant differences.

Table 6 summarizes the 2x3 experimental matrix for the airline pilots.

Table 6 – Airline pilot experimental conditions

Task	Training model		
	Unmodified	Representative	Specific
Low	x	x	x
High	x	x	x

The airline pilots were trained to proficiency in each task, which meant they received two ratings of 3 or better on the stall recovery rating scale (see Appendix). Each task was performed a minimum of three times. After training to proficiency on one of the three models in Table 6, the pilots then assessed a statement using Table 3:

“One or more events in this task had a banking roll-off requiring significant control deflection to counteract.”

After training to proficiency on a task, they were then checked once on that task with the Specific model, which represents the truth case. Then, they assessed the following statement:

“There was no appreciable difference in this checking event as compared to the training events.”

The airline pilots did not know which model they were flying. The ordering of their two tasks was balanced. The session ended with a debrief.

Results

Test Pilot Subjective Evaluation

Of the 13 statements assessed by the test pilots, a statistically significant difference among the four models arose only in one statement. That statement was “One or more events in this set exhibited roll-off requiring significant control deflection to counteract.” The dividing line for statistical significance was $p < 0.05$, or odds less than 1 in 20 that such an observed difference was due to chance. Figure 1 shows the results by stall model broken down by task, as indicated by the altitude name in Table 5 and using the rating scale in Table 3. For all of the figures, the key for the model on the x-axis is as follows:

UM = Unmodified
 UM+R = Unmodified + roll
 REP = Representative
 SP = Specific

The means are shown by symbols. The minimum and maximum are shown by the ends in each of the lines for a given symbol. The 95% confidence interval of the mean is shown by the white space on both sides of the symbol. The standard deviation can be found by multiplying the 95% confidence interval by the square-root of the number of subjects in each independent variable, n , which is 9 in this case. An interaction was found between the model and the task, $F(6,48)=5.01$, $p=0.0013$. This is seen in Figure 1, as the perception of which model required more wheel depended on the task. The amount of roll-off varied, as its magnitude was intentionally randomized, and the interpretation of the word “significant” varied among the pilots. Either factor, or both factors, could be the cause. While some of the ranges in ratings vary from 1 to 6 or from 2 to 7, on average, the Representative model was perceived to have the most roll-off for the low and medium altitude tasks. Interestingly, an instance arose of the Unmodified model receiving a 7 and two 6’s in the low altitude task, even though it had no roll-off.

The assessments for two statements had marginal statistical significance, meaning $0.05 < p < 0.1$. Figure 2 shows one of those statements, which was for “The apparent randomness or non-repeatability of the stall characteristics in this model are similar to the 737-800 airplane,” which an $F(3,24)=3.25$, $p=0.055$. The post-hoc tests revealed that the Unmodified+Roll-off was different from the both the Representative and Specific models at the above marginally statistical significance level.

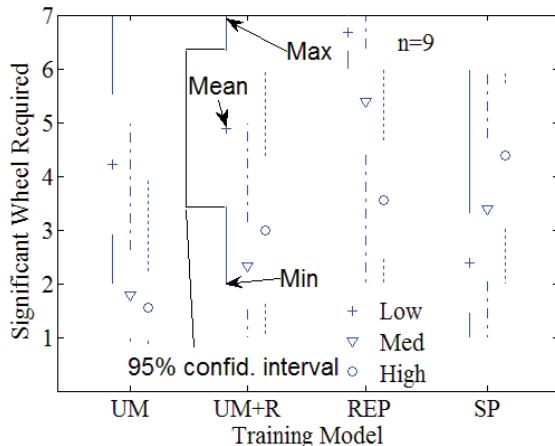


Figure 1 – Perception of roll-off magnitude

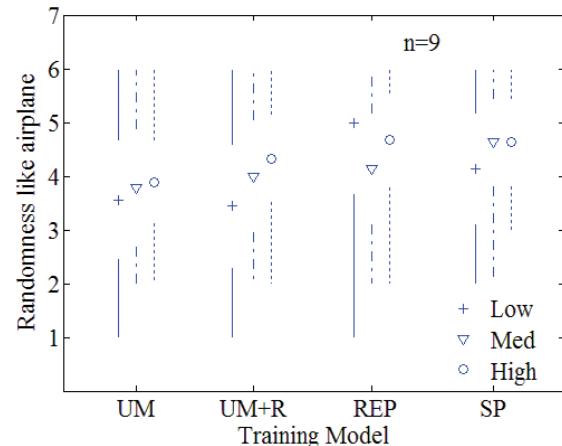


Figure 2 – Perception of randomness

Figure 3 shows the other marginally significant result, which was for the assessment of the statement “The stall buffet in this model is similar to the 737-800 airplane,” which had $F(3,24)=3.02$, $p=0.078$. The post-hoc test showed that the Representative model was different from the Unmodified model. The programming of the buffet was the same for all models, that is, the angle-of-attack response alone determines it. If a pilot went quicker or deeper into a stall with one of the models than for another one, then the perception of the buffet could have been different.

Finally, Figure 4 shows the pilots assessment of the statement “This model could be effectively used to train a student in stall recovery for this task.” Although no statistical differences were detected, the data are shown for this important question. Note that different test pilots expressed wide ranges of opinion for a given model. That is, some rated it a 1 while another rated it a 6 for the same model and task. From the comments, the primary reason for the rating differences was due to the roll-off a model provided. Some test pilots said that, in their experience, the airplane did not roll off as the model they had just assessed and to convey that would be negative training. Other test pilots disagreed and stated the airplane can occasionally roll off as much as the model did. Several pilots said they were not the best person to assess this statement on training, as they were certification test pilots and not trainers.

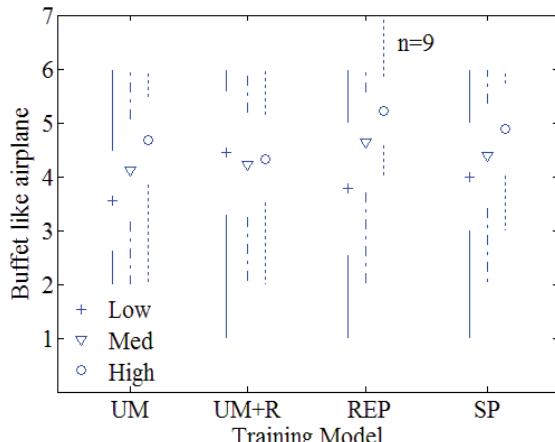


Figure 3 – Buffet assessment

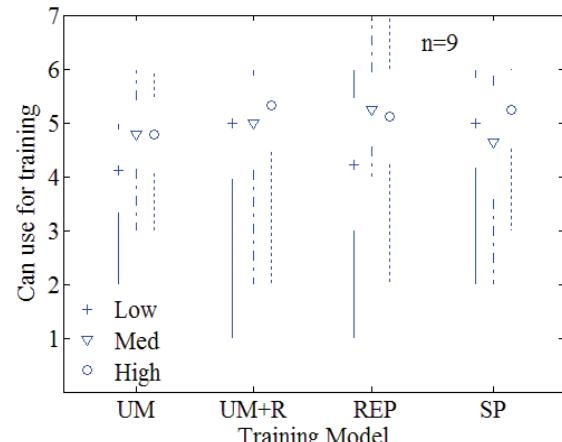


Figure 4 – Training applicability

The test pilots provided comments during the test and debriefs. Below is a selection for each pilot.

Pilot 1: “The buffet needs more intensity and lower frequency.” “Can’t ignore the roll axis. Can’t do it academically. Not sure what you can get out of a generic stall model. A line crew would get more benefit from type-specific training.”

Pilot 2: “The buffet was too light in general. It is more important to mimic the extreme buffet. It was not half as strong as the aircraft. In the aircraft, you cannot read the instruments, and it is disorienting. This buffet was a lot

better than other 737 sims, as there is hardly any buffet in them.” “It would be better to see some roll-off, but I think buffet is more important.”

Pilot 3: “Overall, the buffet model is lacking.” “Initial buffet is the most important and is a great warning.” “I think repeatable sequences are the best way to train.” “I remember working harder controlling roll-off in the airplane than I did in the sim.” “I expected to see more roll-off at high altitudes than I saw in the sim. I saw up to 90 degs of roll in the aircraft at high altitudes.” “I also saw pitch-ups in the aircraft at aft c.g.” “My ordering of importance to get right is buffet, then roll, then pitch, and then yaw.”

Pilot 4: “Buffet fidelity can be a big factor in certain airplanes.” “I think that how far you require the pilot to go before recovery may have a positive or negative effect on the goal of the training program.” “My first thought is that any of the models would be useful.” “Airlines believe there is a cliff at the shaker, and we train them to believe that. I have concerns about removing that cliff mentality.” “I liked the model that challenged me the most. Most of the time the airplane responds well, but sometimes it gets squirrely.” “The turning stalls are harder in the aircraft.” “Doing the proper recovery after pushing over in the high altitude stalls is more important than the initial recovery.” “Need to teach the correct piloting technique for the recovery portion after breaking the stall at high altitude.”

Pilot 5: “Motion cues are required for stall training.” “Some of the models seemed to have rapid roll step inputs at the shaker. I didn’t feel it was indicative of the aircraft.” “The aircraft could roll. I think a student should see that roll in their training.” “The top priority is to show the roll control differences before and after reducing angle of attack.”

Pilot 6: “Some models were reasonably good and some had characteristics inappropriate.” “The model that had a pronounced lateral/directional response is more than what the aircraft has.” “The aircraft does not have more than 10-15 degs of pronounced roll in stalls. It is nothing really significant.” “Aircraft can have deterrent buffet in many cases with flaps up.” “Don’t want to suppress tendency to get into a secondary stall if that characteristic exists.” “The priority should be to accentuate the cues to let them know they are getting close to a stall, then it is a must to show characteristics like the elevator feel shift, and then show that once you break angle-of-attack, the airplane flies better again.”

Pilot 7: “Stall buffet may have been a little weak but satisfactory for sim training.” “A lot of the models could be used.” “For a clean stall, the stick lightening is important to show. A guy on the street may not know it. If he’s been using the same force as in practicing stalls, he could get into a deep stall.” “In training, seeing some extra roll or yawing motion, which could happen, is fine.”

Pilot 8: “The buffet levels will be an eye-opener for pilots who haven’t stalled a jet transport before.” “On the initial demos, my first impression on the buffet was ‘wow, they nailed it.’ Then, as we got into it, I thought it could be improved.” “Some of the roll responses were too pronounced in the sim, some were too small, some about right.” “The predictability in pitch for all models was too predictable.” “This was so much better than any stall training available short of stalling the actual airplane.” “Recognition is so important, so getting the buffet right is the top priority.” “I think too much roll-off is better than not enough. If I’m riding in the back, I would like them to know that roll could be an issue.” “Don’t want the pilot to think the pitch response is really easy, and it isn’t easy to control when you have adrenaline.”

Pilot 9: “It is important for operational pilots to see the high-altitude performance deficit, the elevator feel shift, and buffet.” “Besides the buffet, there isn’t that much feedback from the motion.” “I would resist trying to have a one-size-fits-all stall model.” “It seemed that two models had a random roll, though it might be good for training.” “Had to push on some models in the entry to maintain deceleration rate, so it’s good to see those.” “Priorities are to see the high-altitude performance deficit, then see the strong buffet in the sim so that they are not shocked by it in flight, and then to see the roll.”

Airline Pilot Objective Evaluation

Surprise scenario.

All 45 pilots recovered from the ensuing stall in the surprise scenario with varying degrees of success. Figures 5 and 6 show extremes for one pilot applying the correct stall recovery procedure and for one applying an incorrect procedure, respectively. As the models are proprietary, plots associated with their flight dynamics are intentionally without y-axis numerical values. The Specific model was flown in these two extremes by different First Officers of the same airline.

In Figure 5, the pilot continues to apply a nose-down pitch input with the column until the shaker stops. Control of the roll excursions comes afterwards and does not interfere with the nose-down column input. Though not shown, no pedal was applied, and thrust applications occurred after pitch and roll inputs. The pilot recovered with less than 15 degs of bank and lost 610 ft of altitude.

In Figure 6, with the same y-axis scale as in Figure 5, a different pilot raised the nose after applying full thrust (not shown). He intermittently applied nose-down inputs, but did not do so persistently until the shaker stopped. He entered into a pilot-induced oscillation in the roll axis. Ultimately, the shaker stopped through a combination of slight nose-down pitch and a speed increase from the available thrust at the scenario's low altitude. The shaker remained activated for roughly 3 times longer in the recovery in Figure 6 than that in Figure 5. During the Fig. 6 recovery, the aircraft reached 73 degs of bank and lost 1,700 ft of altitude.

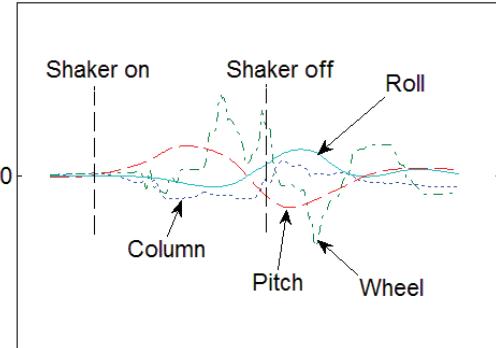


Figure 5 – Stall recovery template applied

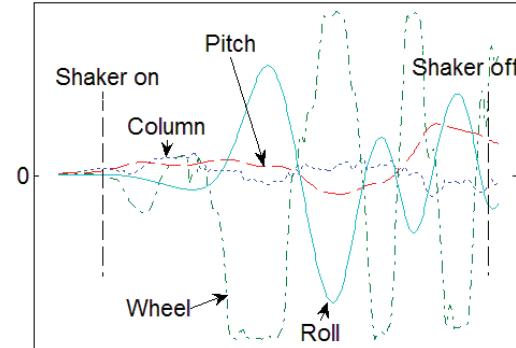


Figure 6 – Stall recovery template not applied

Figure 7 shows whether the stall recovery template of Table 2 was correctly applied in each of the surprise scenarios. To determine this, each record was examined, and all of the following statements had to be satisfied before concluding that the recovery procedure was correctly applied:

- Nose-down pitch control was applied until the shaker stopped
- Wheel input was applied after nose-down control, and it did not take priority over pitch control
- Less than 1 inch of pedal applied while shaker was activated
- Thrust was applied as needed after pitch and roll control, or full thrust had been applied prior to stall warning

The first bullet, application of nose-down pitch control until the shaker stopped, is the key part of the stall template and is easy to measure from the data. The first part of the second bullet is straightforward to measure, but the second part is not. The second part was included as part of the assessment, as situations arose (e.g., Fig. 6) where pilots applied significant, oscillatory wheel inputs while the shaker was activated and should have applied more nose-down input instead. Such actions deviate from the stall template and were noted only when it was obvious.

The third bullet was included, as leaving pedal off the stall template was deliberate during its creation. While 1 inch of pedal is an arbitrary dividing line, it serves as a trigger for unambiguous pedal application. Finally, appropriate thrust application was a condition that needed to be satisfied. The first sign of a problem in this scenario was a precipitous drop in airspeed, which occurred moments before the shaker activated. Many pilots applied thrust in the time between the start of the airspeed drop and the shaker activation. Even though the new stall recovery template is trying to teach pilots not to power out of a stall and instead reduce angle of attack, pilots could have interpreted this event early as a horizontal shear, which it was, and tried to use thrust to counter its effect. If such thrust application occurred in that time period, pilots were not categorized as applying thrust inappropriately. However, they were categorized as doing so if they had not yet applied thrust, the shaker fired, they held pitch attitude, and applied full thrust. In those cases, they were clearly trying to power out of a stall.

Figure 7 shows that, overall, less than one quarter the pilots strictly applied the stall recovery procedure correctly. No difference was detected among the models for this measure. The x-axis shows the model that was flown, which for each pilot was the same as their training model. Figures 8-10 show how often steps of the stall recovery procedure were applied correctly. If any of those steps were applied incorrectly, it led to a 0 for Figure 7. Figure 8 shows that half or less of the pilots, on average, applied nose-down pitch inputs until the shaker stopped. Wide uncertainties existed in the sample means, and no difference was detected among the models for this measure. It is shown, as it is useful to report on how many pilots followed this key step in the recovery procedure when surprised.

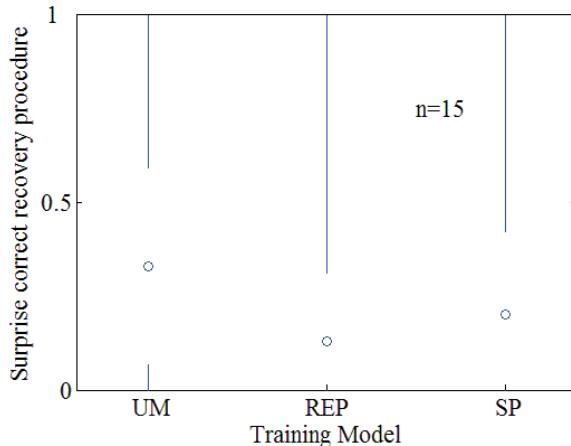


Figure 7 – Recovery procedure correctness

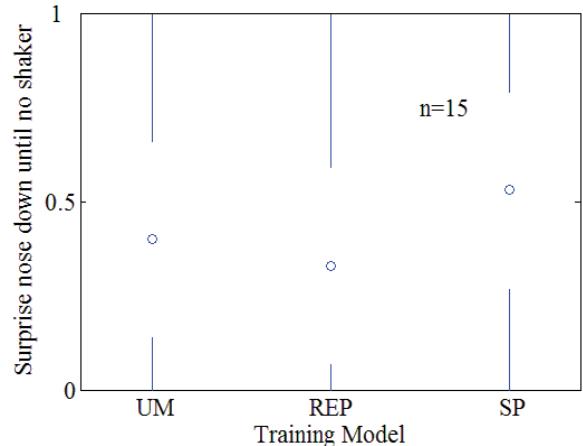


Figure 8 – Appropriate nose down input

Figure 9 shows roughly half of the pilots tried to control the roll axis while the aircraft was stalled. No differences were detected among the models for this measure. Instances arose when pilots created their own difficulties in the roll axis when recovering with the Unmodified model, which does not require much roll control in a wings-level stall recovery.

Figure 10 shows that pilots had a tendency to apply the rudder pedal the most for the Representative model, which did have the largest sideslip response in a stall. Next was the Specific model, which had more sideslip response than the Unmodified model. Approximately one quarter of the pilots still applied pedal with the Unmodified model, even though the stall sideslip response is small for that model. The model had an effect on whether pilots applied pedal inputs, but the confidence for saying so was marginally significant, $F(2,42) = 2.61$, $p=0.085$.

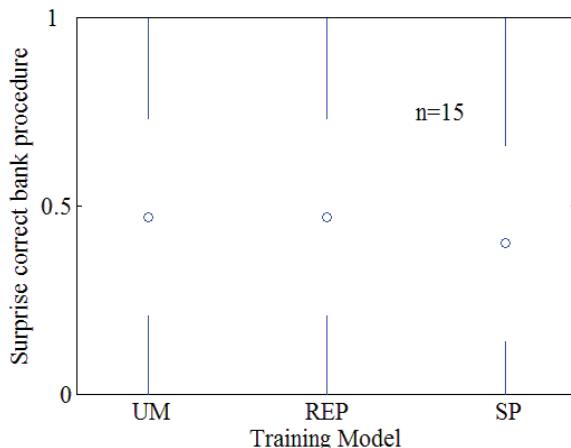


Figure 9 – Appropriate bank procedure

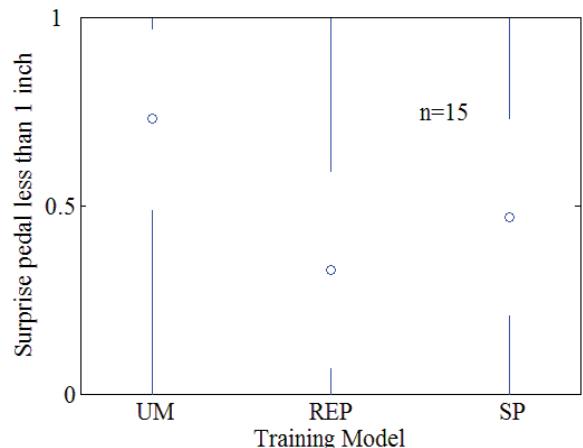


Figure 10 – Pedal usage

To summarize, a total of 10 of the 45 airline pilots applied the stall recovery template perfectly. If the pilot responses are broken down by step in the template, the results were as follows:

- 19 applied nose down pitch control until the shaker stopped
- 20 applied wheel inputs after nose down, and they did not appear to take priority over pitch control
- 23 applied less than 1 inch of pedal when shaker was activated
- 34 applied thrust as needed after pitch and roll control, or applied full thrust prior to stall warning (although this measure potentially did not reflect its possible predominance, as pilots were passed if they had applied full thrust prior to the shaker when seeing the airspeed drop)

Figure 11 shows the maximum roll angle magnitude during the recovery from the surprise. One pilot rolled 90 degs in his recovery with the Specific model. Interestingly, even the Unmodified model had a roll excursion of

nearly 60 degrees, nearly all of which was pilot induced. The model flown had a significant effect on maximum roll angle in the recovery, $F(2,42)=5.04$, $p=0.011$.

Since altitude loss, as long as it is not excessive, should no longer be a factor in stall recovery training, it is with trepidation that Fig. 12 is shown. Although altitude loss is not a consideration for recoveries, for this experiment, it offers a measure of global performance. With both the Unmodified and Representative models, pilots lost an average of approximately 600 ft in this challenging event, which started at 2100 mean sea level (approximately 1950 ft above ground level). On average, more altitude was lost with the Specific model, with one pilot losing nearly 1,700 ft. The model had a significant effect on this measure, $F(2,42)=3.61$, $p=0.036$.

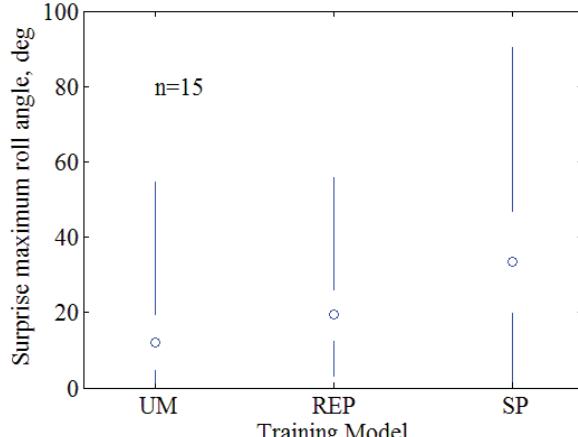


Figure 11 – Surprise roll maximums

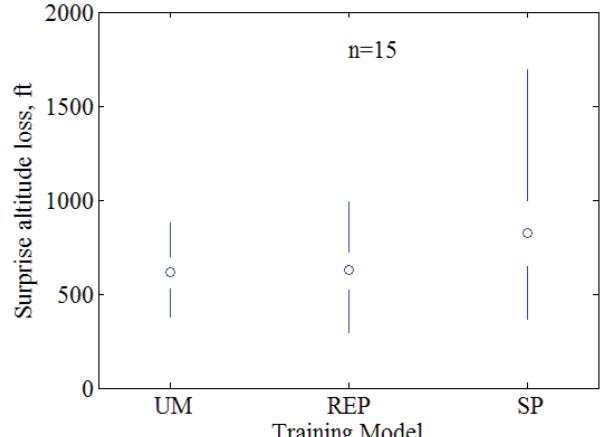


Figure 12 – Altitude loss in surprise scenario

Lastly for the surprise scenario, Figure 13 shows how surprised pilots were by the event. After the surprise, each pilot was asked to assess the statement “I was surprised by that event” using the scale in Table 3. All pilots rated that statement with a 6 or a 7, with thirty-seven pilots rating it “strongly agree” with a 7. Clearly, achieving surprise is simulation is possible.

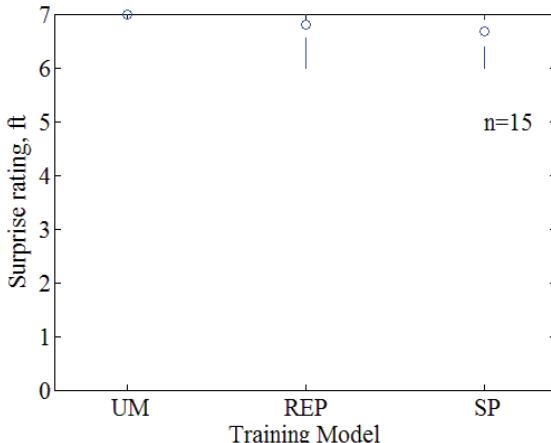


Figure 13 – Pilot subjective rating of surprise

Training and checking.

Figure 14 shows the recovery time in the check as a function of the model in which the pilot was trained. For the low-altitude stalls, a statistically significant effect was found, $F(2,42)=4.01$, $p=0.025$. The difference in the means between the Representative and the Specific model was 2.8 sec. No effect was detected for the high altitude stalls.

Figure 15 depicts the number of times in the check that the pilot tried to control roll before pitch. No statistical differences were detected for the low altitude stalls. A marginally significant effect was detected for the high altitude stalls $F(2,42)=3.14$, $p=0.054$.

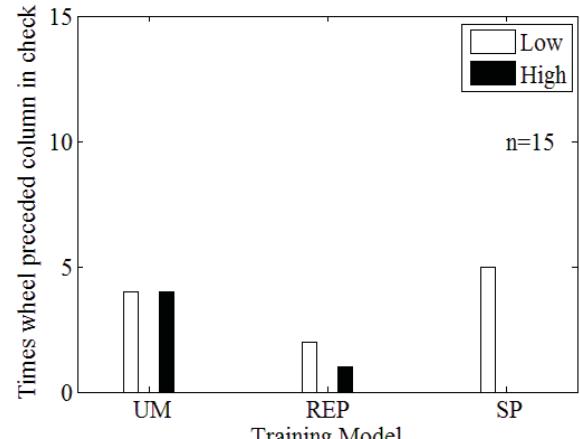
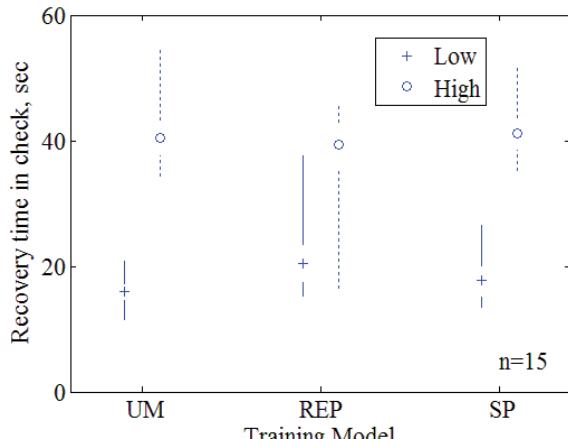


Figure 16 shows the number of secondary stalls at both altitudes during the check as a function of the training model. No statistical differences were found for either the high or low-altitude stalls.

Figure 17 shows the instructor ratings for the recoveries during the check using the scale in the Appendix. No statistical differences were found for either the high or low-altitude stalls. The values of the mean for all check stalls were less than 2, indicating that the average pilot was able to achieve desired performance in the check regardless of the training model. Although the data are not shown, many pilots were not proficient on their first training run regardless of the model, so with instruction, the recovery technique could be learned in these maneuver-based events regardless of the model.

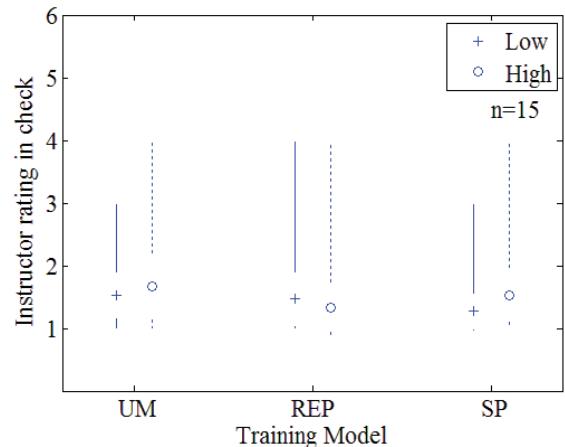
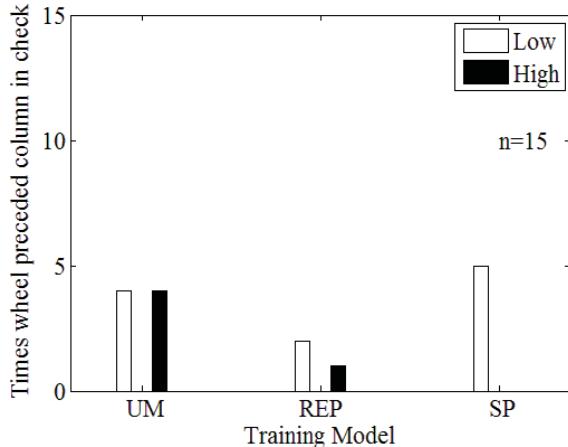


Figure 18 shows the results of the pilots' opinion as to whether significant wheel was required to control a roll-off during any of their training events. No significant differences were detected in these results. As the amount of roll-off was randomized for both the Representative and Specific models, some pilots would have received more roll-off than others, so the wide variance in the results for those two models is expected.

Figure 19 shows the ratings the pilots gave after their checking event with the Specific model. The question asked was if they could tell any appreciable difference between their training events and the checking event. One would expect high ratings, that is, no appreciable difference for the Specific model results, as in those instances, the pilots trained with the same model as they were checked with. However, one difference could still have existed in those cases, as the randomization of the roll-off in that model was on for the training but off (i.e., fixed) for the checking. So, they may have received different levels of roll-off in the training events versus the checking event. Otherwise the models were the same. No statistical differences were detected for the low-altitude stalls, but a statistical difference was detected for the high-altitude stalls, $F(2,42)=5.6$, $p=0.007$. Pilots could tell the difference

between training with the Unmodified model and checking with the Specific model. For this measure, no differences were detected for training with the Representative model and checking with the Specific model; however, the confidence of saying a difference does not exist is equal to the power of the experiment, which has not been determined.

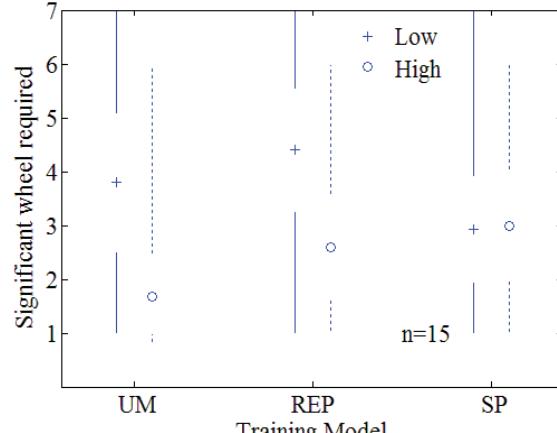


Figure 18 – Wheel requirement in training

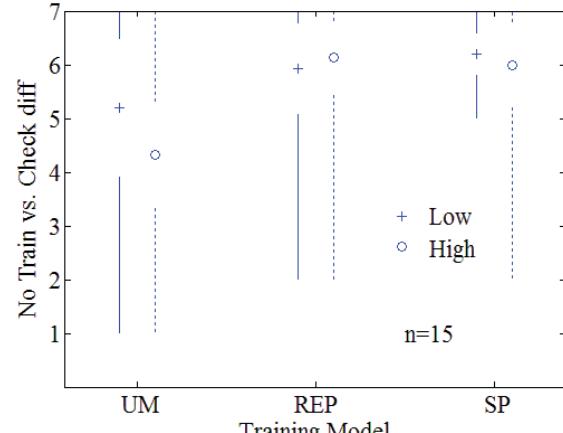


Figure 19 – Training and checking model difference

Discussion

Most all of the test pilots noted the importance of conveying accurate buffet in their comments, and most all thought the buffet model was still lacking, even after doubling its strength. Response averages were indicative that test pilots “somewhat agreed” any of the models evaluated could be used for training. Regarding the roll-off differences, the test pilots perceived the Representative model to have the most roll-off for two of the three tasks, and that the randomness of the roll-off to be more like of the aircraft for the Representative and Specific models. The appropriate magnitude of the roll-off was contested by the test pilots. One pilot thought that the Representative model had excessive roll-off compared to the airplane and could lead to negative training. Another pilot preferred to be challenged by that large roll-off and said that, while the airplane is usually well behaved, such extremes can occur. To show that stall responses are often random in nature, perhaps the best solution is to show what typically can happen and what could happen in the extreme.

Several test pilots noted the value of showing the high-altitude stalls, and one test pilot emphasized that providing guidance on the complete recovery back to stabilized flight was especially important. He was concerned that pilots could create their own problems if not trained properly in such recoveries, and an unnecessary risk could arise in a stall recovery training program if a pilot in the real aircraft misdiagnosed a situation as being a stall when it was not and initiated a poor recovery.

Even though the airline pilots were pre-briefed how to recover completely from a high-altitude stall, problematic high-altitude recoveries did occur. Figure 20 shows the latter portion of a high-altitude stall recovery while attempting to re-establish steady-state flight. This was a pilot’s first training run with the Specific model. Load factor oscillations exceed 1g peak-to-peak. While the peak load factor was within the aircraft’s limit flight envelope, experiencing these load factors for the first time in flight could cause one pilot to ease off on the control to improve the recovery or could cause another pilot to increase his gains further and potentially make matters worse. Without additional instructor tools to feed back load factor after a maneuver in simulation, instructors should monitor the barber pole on the airspeed tape, which effectively shows angle-of-attack oscillations and use this information to educate and improve upon these recoveries.

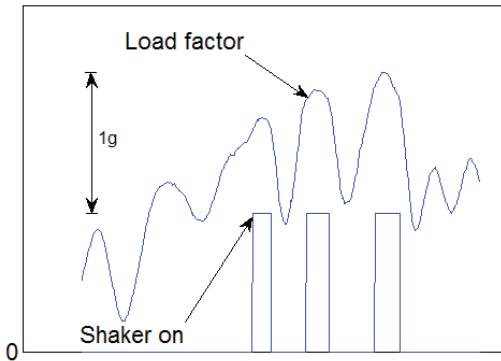


Figure 20 – Load factor in high altitude stall recovery

The results from the surprise scenario turned out to be one of the most valuable parts of the experiment. Instead of the choreographed maneuver training where pilots are told “now we are going to do stall training,” the surprise scenario revealed what might happen if confronted with an unexpected stall in real life. While only 10 of the 45 airline pilots applied the stall recovery template correctly, stringent measures were applied in this assessment for an extremely challenging stall scenario. Some pilots did not want to drop the nose below the horizon to eliminate the stall, and this experience hopefully served to reinforce the fact that an airplane can be stalled at any attitude or airspeed. The fact that most of the pilots strongly agreed that they were surprised was bolstered further by one pilot (the pilot in the Figure 6 recovery) saying “get the nose down, get the nose down” while not getting the nose down during the recovery. Afterwards, he said that he did not know why he knew what he needed to do but could not make himself do it. It is exactly this type of scenario training that is needed in today’s airline training programs.

Recovery times for the low-altitude turning stalls were affected by the model trained with. The average pilot received instructor ratings associated with desired performance during the check regardless of the model in which they were trained. This likely shows the benefit of having an instructor teach the proper method in the training, which has value even for a model that does not have all of the degradations beyond the first indication of a stall (i.e., the Unmodified model, which has some, but not all, of the typical degradations near stall). This result is consistent with the test pilots’ views that training with any of the models would add value.

The issue may be more as to whether having those degradations are valuable for unexpected events in scenario-based training. In those events, a pilot may then be exposed to a situation that is more precarious with the Representative or Specific models than with the more docile Unmodified model. Research has suggested that what is learned in maneuver-based training appears to apply to those situations presented and may not generalize to other unexpected situations.¹⁷ Another issue relates to whether a more difficult model adds more value for a marginal pilot than it does for the average pilot. That is, the average pilot may be insensitive to model differences, but such differences may tease out a training problem with a marginal pilot. In witnessing the variances in technique among the pilots in this study, it may behoove the training community to use models that serve the interests of every pilot, not just the average pilot.

With so much information from a variety of pilots, what is the bottom line? All things considered, either the Representative or the Specific model is closer to what a 737 airplane does in an aerodynamic stall than the Unmodified model, and one or the other appears satisfactory for training stall recoveries. The concerns associated with using the Unmodified model are not as strong as originally believed, at least for the average airline pilot in these tasks. However, the likelihood of exposing deficits in a pilot with less-than-average skills will increase with the use of the Representative or Specific model.

The applicability of using a Representative model in instances when a Specific model is not available appears practical, as the results between those two models were generally similar. In instances when a Specific model can be developed from available flight test data, operators may prefer to use this conventional modeling approach to reduce the risk of negative training. Such assessments of risk should accompany their benefit/cost analysis if they choose to deviate from the conventional approach of creating simulator models from flight data.

Conclusions

Three new stall models were evaluated in a high fidelity simulator using nine test pilots and 45 airline pilots. The principal conclusions were

- It is possible to invoke a strong element of surprise in simulation. Thirty-seven pilots “strongly agreed” and eight “agreed” they were surprised.
- Continued emphasis is warranted in today’s training programs to undo years of applying the incorrect stall recovery procedure. During the surprise scenario, less than one-quarter of the airline pilots strictly followed the correct stall recovery procedure. Less than half applied nose-down pitch control until the stall warning was eliminated.
- Which model used in the surprise scenario had a statistically significant effect on maximum roll angle and altitude loss in the recovery. For both measures, the Unmodified model resulted in less roll excursions and less altitude loss.
- While, on average, the test pilots “somewhat agreed” that any of the models could be used for stall recovery training, wide differences of opinion arose between particular models. Some preferred to be challenged by what could happen in a stall. Others preferred the more benign, typical responses. This is an important training question, that is, do you show what could happen or what happens on average. Perhaps the best answer is to show one of each.
- Even after more than doubling the buffet response, most of the test pilots believed the buffet cue was less in the simulation than in the aircraft.
- Care must be taken to minimize the risk associated with teaching the high-altitude stall recovery in a simulator. The low-g environment of a simulator can mask problematic recoveries when a pilot overcontrols by making abrupt pitch changes at high altitudes.
- The airline pilots were able to train to proficiency with any of the models; some valuable training is likely with any of them, as long as instructors are aware of the deficiencies.
- For the airline pilots, the models did affect the recovery time and whether they applied roll control before pitch control. They were also able to tell a difference between the Unmodified model and the Specific model in the high-altitude stall.
- When flight data are not practically available for stall model development, the alternative approach of using a model based on computational aerodynamics, wind tunnel data, and expert pilot opinion appears feasible.

Acknowledgments

Many people contributed to this experiment. The authors especially thank Sunjoo Advani, Jeard Ballew, Andrew Cheng, Randall Cooper, Joris Field, Bill Gato, Florian Jentsch, Janeen Kochan, George Lyddane, Pete Rohde, Mike Snow, and John Sparkman for their valuable assistance with experimental review, experimental conduct, project management, and computer programming.

References

¹*Qualification, Service, and Use of Crewmembers and Aircraft Dispatchers; Final Rule*, 14 CFR Part 121, Federal Aviation Administration, Federal Register, Vol 78, No. 218, Nov. 12, 2013, <http://www.gpo.gov/fdsys/pkg/FR-2013-11-12/pdf/2013-26845.pdf>

²Code of Federal Regulations, Title 14, Part 60, Flight Simulation Training Device Initial and Continuing Qualification and Use, Washington, DC.

³ICAO 9625, “Manual of Criteria for the Qualification of Flight Simulation Training Devices,” Doc 9625, Third Edition, International Civil Aviation Organization, Montreal, Quebec, Canada, 2009.

⁴Advani, S.K., Schroeder, J.A., and Burks, B., “What Really Can Be Done in Simulation to Improve Upset Training?” AIAA Modeling and Simulation Technologies Conference, Toronto, Canada, Aug. 2010.

⁵Gingras, D.R., Ralston, J.N., and Wilkening, C.A., “Improvement of Stall Regime Aerodynamics Modeling for Aircraft Training Simulations,” AIAA Guidance, Navigation, and Control Conference, Toronto, Canada, Aug. 2010.

⁶Schroeder, J.A., “Research and Technology in Support of upset Prevention and Recovery Training,” AIAA Modeling and Simulation Technologies Conference, Minneapolis, MN, Aug. 2012.

⁷Field, J. and Shikany, D., “Simulator Fidelity Requirements for Upset Prevention & Recovery Training,” 8th International Flight Crew Training Conference, The Royal Aeronautical Society, London, England, Sept. 2013.

⁸Foster, J.V., Cunningham, K., Fremaux, C.M., Shah, G.H., Stewart, E.C., Rivers, R.A., Wilborn, J.E., and Gato, W., "Dynamics Modeling and Simulation of Large Transport Airplanes in Upset Conditions," AIAA Guidance, Navigation, and Control Conference and Exhibit, San Francisco, CA, Aug. 2005.

⁹Field, J., Roza, M., and Smaili, H., "Developing Upset Cueing for Conventional Flight Simulators," AIAA Modeling and Simulation Technologies Conference, Minneapolis, MN, Aug. 2012.

¹⁰Groen, E., Ledegang, W., Field, J., Smaili, H., Roza, M., Fucke, L., Nooij, S., Goman, M., Mayrhofer, M., Zaichik, L., Grigoryev, M., and Biryukov, V., "SUPRA – Enhanced Upset Recovery Simulation," AIAA Modeling and Simulation Technologies Conference, Minneapolis, MN, Aug. 2012.

¹¹Zaichik, L.E., Yashin, Y.P., and Desyatnik, P.A., "Some Aspects of Upset Recovering Simulation On Hexapod Simulators," AIAA Modeling and Simulation Technologies Conference, Minneapolis, MN, Aug. 2012.

¹²Ko, S.K. and Grant, P.R., "Development and Testing of an Adaptive Motion Drive Algorithm for Upset Recovery Training," AIAA Modeling and Simulation Technologies Conference, Minneapolis, MN, Aug. 2012.

¹³Donaldson, S., Priest, J., Cunningham, K., and Foster, J.V., "Upset Simulation and Training Initiatives for U.S. Navy Commercial Derived Aircraft," AIAA Modeling and Simulation Technologies Conference, Minneapolis, MN, Aug. 2012.

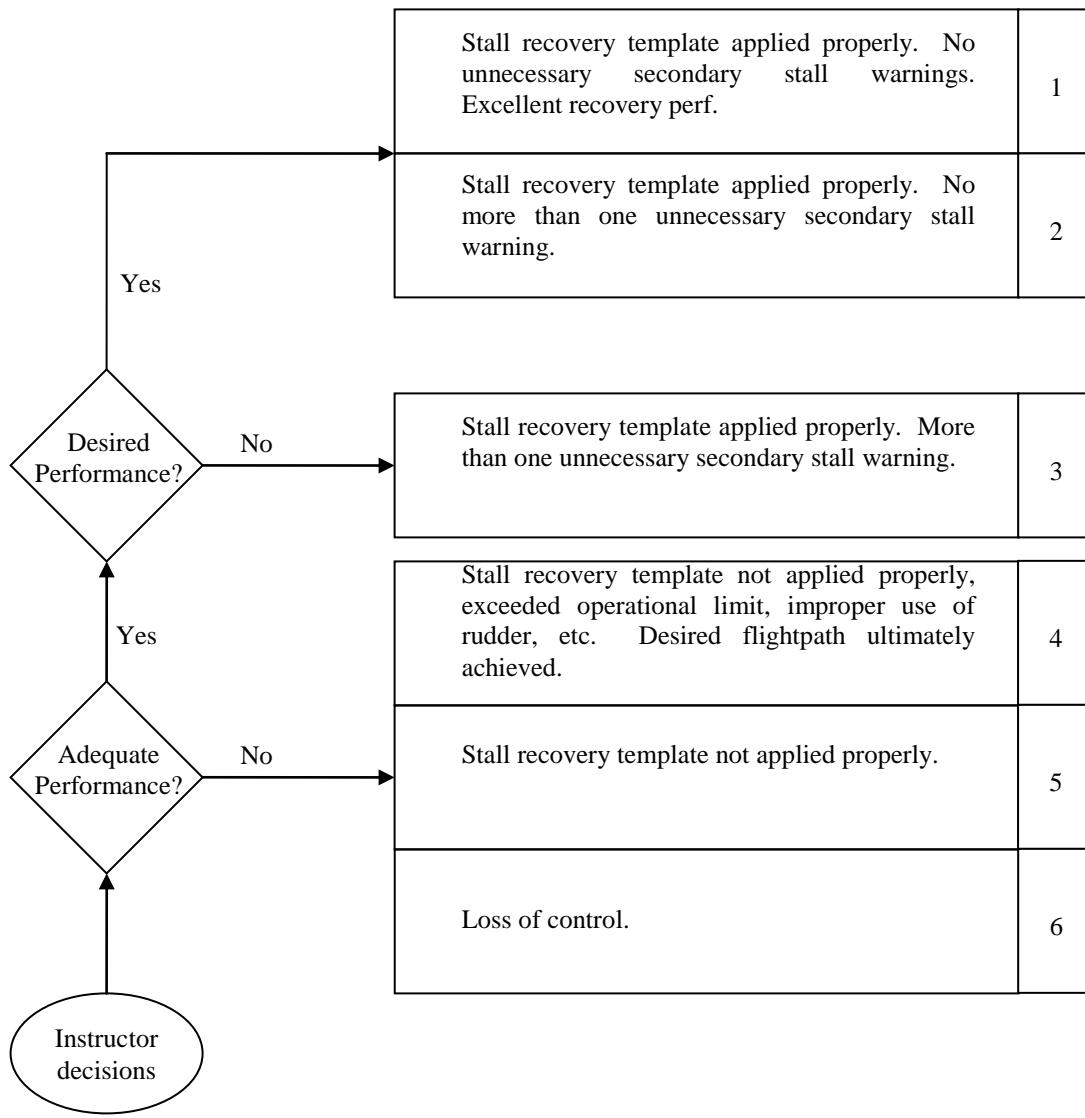
¹⁴Federal Aviation Administration, "Stall and Stick Pusher Training," Advisory Circular No. 120-109, Washington DC, Aug. 2012.

¹⁵Federal Aviation Administration, "Airline Transport Pilot and Aircraft Type Rating Practical Test Standards for Airplane," FAA-S-8081-5F (with Changes 1, 2, 3, 4, 5, & 6), Washington DC, July 2008.

¹⁶Cooper, G.E. and Harper, R.P., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," NASA TN D-5153, National Aeronautics and Space Administration, Washington, DC, Apr. 1969.

¹⁷Casner, S.M., et. al., "The Effectiveness of Airline Pilot Training for Abnormal Events," *Human Factors*, Vol. 55, No. 3, June 2013, pp. 477-485.

Appendix
Stall Recovery Rating Scale



Desired performance

Meets adequate performance criteria AND:

No more than one reoccurrence of stall warning (i.e., stick shaker, buffet, aural warning)

Adequate performance (each must be satisfied) (extracted from Ref. [FAA, 2008])

Prompt execution of stall recovery template (i.e., correct order, no secondary full stalls)

Recover in a safe and timely manner

Recover to maneuvering speed and flight path appropriate for the airplane's configuration

Recover without exceeding the airplane's limitations

Recover without losing excessive altitude consistent with the airplane's performance capabilities

Demonstrates smooth, positive control during the recovery