DOT/FAA/AM-21/10
Office of Aerospace Medicine
Washington, DC 20591

3-D Safety Margin Profiles when Using Head-Up Display (HUD) for Takeoff in Low Visibility and High Crosswind Conditions

Daniela Kratchounova[1]
Inchul Choi[2]
Theodore Mofle[2]
Larry Miller[3]
Jeremy Hesselroth[3]
Scott Stevenson [3]
Mark Humphreys[2]

[1] Federal Aviation Administration
Civil Aerospace Medical Institute,
Oklahoma City, OK, USA
Daniela.Kratchounova@faa.gov

[2] Cherokee Nation 3S, Tulsa, OK, USA
Inchul.CTR.Choi@faa.gov,
Theodore.C-CTR.Mofle@faa.gov, Mark.Humphreys3@gmail.com

[3] Federal Aviation Administration
Flight Technologies and Procedures Division,
Oklahoma City, OK, USA
Larry.C.Miller@faa.gov, Jeremy.J.Hesselroth@faa.gov,
Scott.Stevenson@faa.gov

1. Report No. DOT/FAA/AM-21/10	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle 3-D Safety Margin Profiles when Using Head-Up Display (HUD) for Takeoff in Low Visibility and High Crosswind Conditions		5. Report Date March 2021
		6. Performing Organization Code
7. Author(s) Kratchounova, D.[1], Choi, I. [2], Mofle, T. [2], Miller, L. [3], Hesselroth, J. [3], Stevenson, S. [3], Humphreys, M.[2]		8. Performing Organization Report No.
9. Performing Organization Name and Address [1] Federal Aviation Administration Civil Aerospace Medical Institute, Oklahoma City, OK, USA Daniela.Kratchounova@faa.gov [2] Cherokee Nation 3S, Tulsa, OK, USA Inchul.CTR.Choi@faa.gov, Theodore.C-CTR.Mofle@faa.gov, Mark.Humphreys3@gmail.com [3] Federal Aviation Administration Flight Technologies and Procedures Division, Oklahoma City, OK, USA Larry.C.Miller@faa.gov, Jeremy.J.Hesselroth@faa.gov, Scott.Stevenson@faa.gov		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplemental Notes		
16. Abstract The role of crosswinds in the relationship between crew workload and size of safety margin during low visibility takeoffs using HUD was examined. 3-dimensional (3-D) safety margin profiles for different low visibility takeoff conditions and guidance types were generated utilizing the raw and standardized scores for the six NASA-TLX subscales, with and without the effect of crosswinds. The results underlined the critical role for aviation safety of a) building an evocative shared mental model within the pilots' community of the multi-faceted nature of safety margin; and b) having a clear understanding of the complex interplay between the many factors that affect its shape and size.		
17. Key Words Safety Margin, Crew Workload, Head-up Display (HUD), NASA-TLX, Low Visibility Operations	18. Distribution Statement Document is available to the public through the Internet: (http://www.faa.gov/go/oamtechreports/)	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	26	

Introduction

In a previous research effort [7], we identified the differential effects of head-up display (HUD), HUD with localizer guidance symbology, runway visual range (RVR) and runway centerline lighting infrastructure on crew workload as measured by the NASA Task Load Index (NASA-TLX) scores during low visibility takeoff operations. Only the total weighted NASA-TLX scores were used for those analyses and crosswinds were not a factor of primary interest.

Here, we analyzed the role of crosswinds in the relationship between crew workload and safety margin as measured by the raw NASA-TLX subscale scores. The information gained from these analyses enabled the creation of 3-dimensional (3-D) safety margin profiles across different low visibility takeoff conditions and guidance types. Specifically, the scores for the six NASA-TLX subscales were used to plot each safety margin profile with and without the effect of crosswinds. In addition, we discuss the importance of pilots recognizing that handling crosswinds safely requires awareness not only of the aircraft and their own personal limitations, but all factors that could directly, or indirectly, affect the size of safety margin.

The notion of margin of safety is fundamental to the notion of aviation itself. Safety margins apply to many areas of flight operations including flight environment (e.g., weather), ground infrastructure (e.g., runway lighting), etc. Furthermore, aircraft design involves multiple layers of safety margins intended to improve safety without unnecessarily limiting aircraft and human performance. Numerous factors affect these safety margin layers and may dynamically change the size of each layer in the different phases of flight, and across varying operational and environmental conditions.

Conceptually, the size of a single safety margin layer could be defined as the “distance” between a crew workload profile, and a potential incident or accident boundary in a given flight situation. In this context, the probability of safety margin reduction as a function of some type of hazard, or a combination of hazards, is referred to as risk. Besides, the way people recognize risk is inherently subjective and reflects their: a) previous experience with a particular hazard; b) perception of the potential, direct or indirect, negative consequences and how imminent these consequences are; c) sense of control over the situation; and d) individual biases toward competency and control. Therefore, building strong mental models about how the effects of one factor (e.g., crosswinds) interact with the effects of other operational and environmental factors, and new technologies; is critical for the safety of flight.

Background

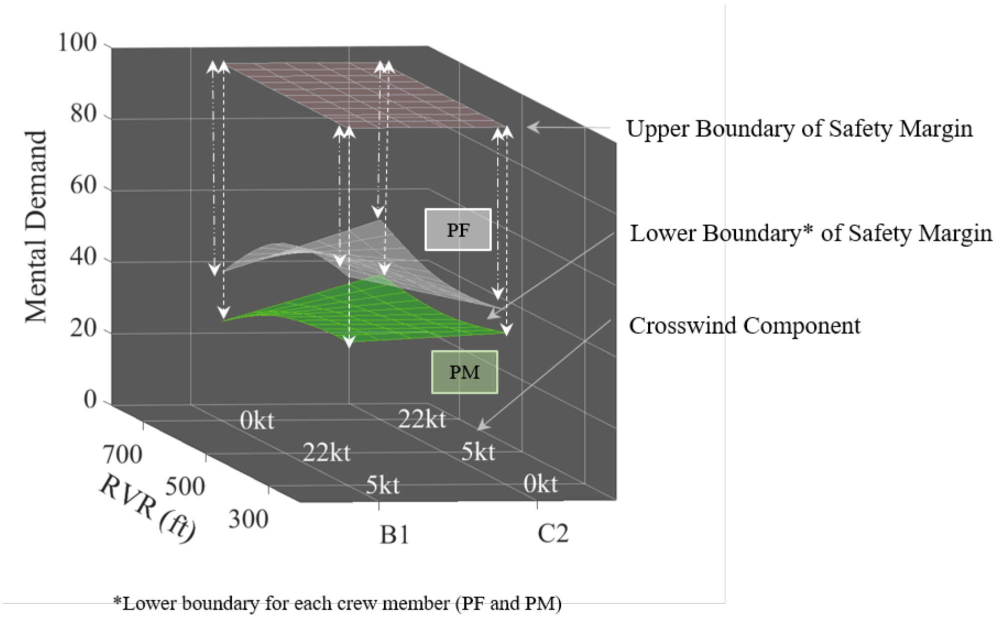
In aviation, as with other high-risk operational environments, there is a constantly fluctuating margin between two distinct workload boundaries. That is, a lower boundary that could represent a pilot/crew's current workload level resulting from performing normal pilot tasks and responses to hazards presented by actual conditions; and an upper boundary representing a pilot/crew's total capacity to positively respond to hazards and safely manage tasks under those conditions.

Hart [5] echoes the notion that workload is "the human cost (e.g., fatigue, stress, illness, and accidents) of maintaining performance" (p. 904). When that cost is unacceptably high, the capacity of a human operator to perform a given task safely may be depleted. As the distance between the two boundaries decreases, the safety margin decreases, leaving less spare capacity for the pilot to resolve hazards or successfully complete required tasks.

For the purposes of this research, the raw scores on the six NASA-TLX subscales established the lower boundary. The upper boundary was "mapped" to the upper limit of the subscales [6]. The highest score on the NASA-TLX workload scale is 100, signifying the upper limit of pilots' total capacity to manage tasks safely under those conditions [4].

Fig. 1 portrays an example of NASA-TLX Mental Demand profiles for the pilot flying (PF) and the pilot monitoring (PM) across two types of takeoff guidance, three levels of RVR, and various crosswind components. The PF and PM perform separate duties on the flight deck and therefore are considered as different populations. Thus, a crew workload profile includes the profiles for the PF and the PM.

On the NASA-TLX Mental Demand subscale, the PF reported experiencing higher workload than the PM for the given set of conditions (Fig. 1). Furthermore, the differences between the reported workload levels for the PF and PM also varied across conditions. Consequently, the resulting size of safety margin for each crewmember was different.



Legend:
B1 – Baseline 1: HUD, no LOC guidance, runway centerline marking only
C2 – Condition 2: HUD with LOC guidance, no runway centerline marking, no centerline lights

Fig. 1. Safety margin's lower boundary for each crewmember on the Mental Demand NASA-TLX subscale.

While the number of actual conditions that could affect safety margin is seemingly unlimited, the impact of crosswinds on takeoff and landing provides a rich example to consider. For high crosswind conditions, the lower boundary could be defined as the workload associated with resolving the aircraft directional control hazard caused by the crosswind component. In this particular case, the upper boundary limit could be determined as the pilot's total capacity to control the aircraft safely under crosswind conditions.

It is important to note that the controllability of the aircraft is not solely based on the all-engines operating case. The crosswind considerations must also include engine failure considerations. Under strong crosswind conditions, the aircraft naturally tends to weathervane into the wind due to side forces exerted on the aircraft fuselage. Additionally, an engine failure in a twin-engine aircraft causes the aircraft to yaw in the direction of the failed engine. The torque generated from the thrust differential makes an upwind engine failure under high crosswind conditions the most critical engine failure case for controllability. If the actual runway crosswind conditions remain below the limits of both the aircraft and the pilot, then the safety margin is preserved. However, if the crosswinds exceed the capabilities of either the pilot or the aircraft, and/or are combined with other risk factors such as inclement weather, then the safety margin may be compromised resulting in an accident.

Nevertheless, safeguards and mitigating factors that positively affect the lower boundary level could preserve or enhance the safety margin. For example, in high crosswind operations without the use of advanced flight deck technologies such as a HUD, the lower boundary workload level could be optimized by either utilizing a runway direction more closely aligned with the wind, or simply, waiting for the winds to subside. Similarly, factors such as training and experience, crew resource management (CRM), and new technologies may have a profound impact, as well.

Training and Experience

One potentially large contributor to increasing the safety margin in a crosswind environment would be optimizing the lower boundary limit by increasing pilots' skill level through standardized initial and recurrent training. This includes learning about landing techniques such as de-crabbing[1] the aircraft just prior to touchdown, and building a mental model of the correct sight picture for wings level and on runway centerline, both in the daytime and at night. It also includes rigorous training on what actions are to be taken if the aircraft's performance is outside of specific parameters. For example, an aircraft operator could establish a training standard specifying that if a safe landing cannot be achieved with wings level and within 20ft of centerline, a missed approach must be executed. The "must" is important as it relates to safety margin because it relieves the pilot of having to quantify the size of the safety margin in a highly dynamic and time-compressed environment.

Correctly determining where the lower boundary actually resides with respect to the upper boundary is a nearly impossible task in real time. However, pilots can easily determine when they are approaching the limits of a specific training standard that is tailored to ensure that as much of the safety margin is preserved as possible. If there were a real or perceived pressure not to execute a missed approach and just land the plane regardless of its position relative to the prescribed one, the safety margin may be reduced.

Through the process of continuous training, a culture of safety is created. More, a well-established safety culture gives pilots the agency and confidence to "do the right thing" every

[1] "Crabbing" is to point the nose of the plane into the wind. The plane flies sideways, similar to how a crab walks.

time. Building upon the knowledge and skills gained through training, real-world experience and the practical application of those skills is often the best instructor. This encompasses not only learning directly from personal experiences, but also learning indirectly by observing the experiences of other pilots. For example, a pilot may observe the full effects of crosswind while lined up for takeoff during high crosswind conditions, as the plane ahead rolls off for takeoff. If the preceding aircraft fails to apply sufficient crosswind controls into the wind, the upwind wing would rise and the engines on the downwind wing would come very close to the ground. As a result, the aircraft may experience a wing rock ensuing in a roll angle with the upwind wing low, before returning to a wings-level condition. The pilot(s) in the waiting aircraft would have no way of knowing all of the factors that contributed to this outcome. However, they could gain valuable insight by observation and in the future, recognize the likely signs of insufficient crosswind controls into the wind followed by a sharp overcorrection in response to the wing rise. Ultimately, retaining and reinforcing knowledge of proper crosswind control techniques improves the safety margin by optimizing the workload profile that defines its lower boundary.

Crew Resource Management (CRM)

In multi-crew aircraft, CRM has a significant impact on safety margin. Through crew briefings and other activities, CRM affects the lower boundary by creating a shared mental model that serves to quantify, organize, and articulate all of the external factors that the aircrew is managing during a particular phase of flight. This organization of tasks optimizes pilots' cognitive workload, therefore preserving the safety margin. At the same time, by delegating and sharing tasks, pilots are effectively able to redistribute their individual workload level where an optimized crew workload level reflects the pilots working together performing normal pilot tasks and resolving hazards. With regard to crosswind takeoffs and landings, CRM helps both pilots have a shared mental model about how the winds are likely to affect the aircraft, what the acceptable takeoff and landing parameters are, and what actions should be taken if those parameters are exceeded. As a result, the size of the safety margin is maximized, as well.

New Technologies

New technologies on the flight deck are intended to enhance the safety margin, improve performance, and optimize crew workload. At the same time, such technologies may initially raise the lower boundary and consequently reduce the safety margin due to elevated cognitive workload levels associated with the lack of experience in flying with these technologies. For example, in the case of crosswind takeoffs and landings using a HUD, there are three main tools used: the flight path vector (FPV), the boresight symbol, and the wind readout. The FPV and boresight symbol are very useful, especially at night and in low visibility conditions, because they provide the pilot with a visual cue of the angular difference between the actual flight path and the longitudinal axis of the aircraft when no external visual cues are available. The wind readout is useful during high crosswind landings because the winds are seldom steady throughout the approach and landing.

As the wind velocity and vector continually change, this requires appropriate adjustments in flight control inputs. Without any indication depicting the general trends in wind velocity and direction, the flight control inputs may lag in response to the changing crosswind conditions. The wind readout helps reduce the lag time and, when used in conjunction with the FPV, it helps refine the required flight control inputs. Takeoff operations similarly benefit from the HUD. From immediately after takeoff and throughout the departure climb segments, the FPV compared to the boresight symbol provides a near instantaneous visual depiction of the actual aircraft flight path compared to the aircraft's longitudinal axis. By placing the FPV over the top of the flight director cue, and cross-referencing the wind readout, a pilot is able to reduce the wind drift and tracking error relative to the planned departure track when compared to operations without the assistance of the HUD symbology.

The combination of FPV, boresight symbol, and wind readout on the HUD enhances pilot performance and optimizes workload levels because the information is immediately actionable for proper application of control inputs. This eliminates the typical lag associated with waiting for a deviation to occur, for that deviation to be recognized, and then for corrective inputs to be applied to return the aircraft to the planned departure track. Once a pilot overcomes the initial challenge to incorporate this new information into their cognitive process, the use of a HUD positively affects the lower limit by providing the pilot with highly actionable information that reduces pilot response time, thereby enhancing the safety margin.

In summary, and for the purposes of this study, a safety margin is the relative spread between the lower boundary, representing a pilot's current workload level, and the upper boundary limit, representing a pilot's total capacity to handle normal pilot tasks and resolve hazards. The lower boundary limit is profoundly influenced by training and experience, CRM, and the introduction of new technologies on the flight deck. New techniques already exist (e.g., use of a HUD) that are capable of further optimizing the safety margin by limiting the lower boundary to a certain level to ensure it remains well below the upper boundary during all phases of flight, types of operations and environmental conditions.

Personal Crosswind Limits

On any given day, a pilot's performance is not limited solely by the extent of their training. For decades, the literature has extensively documented factors that potentially negatively affect pilot's performance, including factors such as fatigue, use of over-the-counter medications, alcohol consumption, and other stressors in their personal lives that detract from their ability to focus. It is an established practice that pilots need to assess their individual fitness to perform flight duties before every flight. While that is true for both commercial air carriers and general aviation, the methods of addressing those impacts may differ between the two communities.

Operational limitations for commercial air carriers are very specific, and typically could be found in the Aircraft Flight Manual (AFM), the company's Flight Operations Manual, or the Operations Specifications that govern a carrier. For any set of operational conditions, a pilot is authorized to operate only within the specified limits. Pilots in this community are trained and evaluated for operations according to predefined standards. A pilot is expected to report for duty fully prepared to operate at that level. If any personal factors exist that would prevent them from safely operating up to the standard, it is customary to decline the flight assignment so another pilot is assigned to the flight. This makes the fitness for duty decision a binary choice. For the community of commercial air carrier pilots, this eliminates the need for setting personal minimums based upon other operational or environmental factors.

In contrast, general aviation has fewer prescribed restrictions imposed on the pilots. This places more responsibility on pilots to assess not only their personal fitness for flying duties to determine a set of personal minimums, as well. General aviation pilots may set their personal minimums according to the aircraft limitations in most situations. However, in certain situations the pilot may choose to set a lower personal minimum. In a stark contrast to commercial air carriers, if a limit is not in the AFM then there may be a few situations where a general aviation pilot would set personal minimums based on the Aeronautical Decision-Making (ADM) guidance [2]. Such circumstances might involve weather, runway conditions, location, and in some cases, a known design issue inherent to a certain airframe that pilots learn to pay special attention to, during training.

One situation where a pilot may consider setting a personal crosswind limit lower than the ones set forth by the AFM would be during low visibility takeoff. Under FAR Part 91, which is the sole governing regulation for most general aviation operations, there are no takeoff minimums. Therefore, when the visibility is reduced significantly, pilots may choose not to attempt a takeoff in crosswinds that are close to, or at the limit of the aircraft. For example, a particular aircraft may have a “maximum demonstrated” crosswind of 28kt for takeoff. This may not be an aircraft limitation per se, but simply the maximum crosswind the test pilots demonstrated during certification [8]. In these situations, pilots often choose to set their personal limit to a value less than the maximum demonstrated crosswind. While 28kt is a very high crosswind, there are techniques that would help pilots mitigate potential hazards when operating the aircraft close to these demonstrated conditions. For example, one such technique is the use of normal crosswind control inputs then slowly reducing the amount of input as ground/airspeed increases the effectiveness of the control surfaces. A different technique to consider is a slightly higher rotation speed and/or a faster rotation especially if the crosswind is gusting. The intent is to keep the wheels in contact with the ground as long as possible and rotate the aircraft away from the ground as quickly as possible.

In degraded visual conditions, the number of outside visual references are reduced substantially. Therefore, pilots may consider lowering the crosswind they are willing to take off with, and do so relative to the amount of reduced visibility. Following with the example above, if the visibility is low enough; pilots may reduce the crosswind they are willing to accept from 28kt to 20kt or even lower, depending on the specific conditions.

Another factor in considering a personal crosswind limit is the runway condition. Contingent on the amount of contamination (e.g., water, ice) on the runway, pilots should consider reducing the amount of crosswind they are willing to accept due to reduced traction of the tires on the pavement. In some AFMs and countries of registry, this is an actual limitation and aircraft registered there would be limited to 28kt for wet runways with “good” braking reported, 15kt for “average”, 10kt for “poor” and for a coefficient of friction less than 0.3, the takeoff would be prohibited. Similarly, for “contaminated” runways, the aircraft would be limited to 10kt crosswind and prohibited with the coefficient of friction less than 0.3. For most countries however, there are no limitations, therefore it becomes the responsibility of the Pilot-In-Command to establish a crosswind limit suitable to the runway conditions. If a limitation is not introduced during training, it is the pilot’s responsibility to draw from experience to establish acceptable and safe personal minimums. Flying with more senior pilots may help with establishing such minimums. Essentially, they may be passed down to junior pilots in an informal manner.

In addition, certain aircraft design characteristics might make the pilot more susceptible to error during crosswind takeoffs and landings therefore warranting personal limits that are more conservative. Aircraft with narrowly spaced main gear, a low belly and wings, and highly swept wings may be good examples. Moreover, defined crosswind limitations or “demonstrated” crosswinds may not exist for some aircraft [8]. During initial training, pilots may receive some suggested guidance on acceptable crosswind limits. However, these could vary greatly across flight crews since there are no demonstrated or limiting crosswinds defined by the manufacturer. As a result, crews may report an unexpected and sometimes unnoticed issue of wing “scrape” during high crosswind and excessive pitch takeoffs and landings. A “word-of-mouth” pitch/bank combination could be what crews set when flying such airframes.

Location, unique terrain features, and weather phenomena associated with that location also warrant setting personal crosswind limits lower than what is normal or defined in the AFM. Due to the runway direction and layout of the terrain, when there is a direct crosswind, nearby mountains may create a Venturi effect. These phenomena add up to winds that are significantly stronger at 50-100ft above the runway than is reported by the automated weather observation service. Consequently, pilots should expect that shortly after rotation while still at low airspeeds, they might encounter moderate to severe turbulence that could potentially affect the controllability of the aircraft. Given these factors, and when winds at the airport are reported to be higher than 10-15kt of direct crosswind, pilots should plan delaying the departure until winds subside or change direction. Through experience of known weather and terrain phenomena, this is one instance where it may be prudent to set up lower crosswind limits than what would otherwise be considered typical.

Crosswind Takeoff Considerations

Regardless of whether a pilot is performing flight duties as an air carrier operator or as a general aviation pilot, they likewise require the ability to determine the actual crosswind component of the wind relative to the runway heading. Fortunately, the actual runway crosswind conditions are easy to estimate based on the winds reported from the tower controller. While the Automatic Terminal Information Service (ATIS) broadcasts report the average wind velocity and direction valid at the time of the ATIS recording, for takeoff purposes they are mostly useful only for runway planning and performance data calculations. At tower-controlled airfields, the controller typically provides current runway wind conditions as part of the takeoff clearance to civil aircraft. At those same airfields, it is mandatory for the controller to provide runway wind direction and velocity to military aircraft because crosswinds are specifically factored into performance data for aircraft that are certified under military specification. Using the tower-reported wind direction and velocity, a pilot can calculate the crosswind component of runway wind by analyzing the horizontal component of the wind vector relative to the runway heading. For operations at airports without an operating control tower, an Automated Weather Observation System (AWOS) may be available that broadcasts nearly real-time wind data that can be used for performance calculations.

Crew Workload Considerations

To achieve optimal system performance, system designers need to take into consideration the overall operator workload at all stages of system design and operation. The NASA-TLX is a human-centered rating scale in which information about the size and sources of six dimensions of workload are combined to develop a sensitive and reliable subjective assessment of workload [6]. It was developed based on the assumption that a combination of six workload related factors (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) represent the workload experienced by most people performing most tasks. These dimensions were selected after extensive analyses of factors that identify the subjective experience of workload for different people performing activities ranging from simple to complex tasks such as flying an aircraft [4, 6].

According to Hart [5], the majority of studies that used NASA-TLX addressed a question about the user interface design, and 31% of them focused on display design. Furthermore, the author reported that a common variation of the scale is to conduct subscale-rating analyses instead of generating a single overall workload score. Over 40 studies used this approach and demonstrated the potency of the scale and the diagnostic value of the component subscales. The author concluded that the high reliability, sensitivity, and utility of the NASA-TLX component ratings allow for a very narrow identification of sources of a workload or performance problems [5].

Our original research [7] included analyses that used only the total weighted NASA-TLX scores. With this follow-on research, we focused on identifying the size of the safety margin based on the raw scores from each of the NASA-TLX subscales. Specifically, we looked for an insight to whether or not the introduction of the HUD, with either just its basic HUD symbology or with the additional localizer takeoff guidance, contributed to a more optimized crew workload profile and helped preserve the safety margin under low visibility conditions. We hypothesized that any contribution to a more optimized pilot workload profile due to the use of a HUD, and a dedicated set of takeoff guidance symbology, could also allow for a larger takeoff safety margin in higher crosswinds conditions.

During each crew's briefing session, we found that it was easier for the pilot evaluators to look at the six NASA-TLX subscales as two distinct groups. One of the groups included the first three subscales - all pertaining to the task itself. Specifically, we requested that the pilots assess the mental, physical, and temporal demands as sources of workload as related to the nature of the task in the specific conditions it was performed. The second group consisted of the last three subscales, all associated with the person performing the task. In this case, the pilots were asked to rate their individual performance on the task, the total level of effort they felt was required to meet the demands of the task and achieve that performance; and the level of frustration they experienced while performing the task in each set of conditions.

Based on the raw NASA-TLX subscale scores, with this follow-on study, we introduced a novel method of visualizing the data in 3-D by generating safety margin profiles across the different guidance types, RVR values, crosswinds, for normal and abnormal flight operations; and with and without the use of a HUD for takeoff. By generating safety margin profiles with the raw and standardized scores for each NASA-TLX subscale, the objective was to help the pilot community build an evocative mental model of the effect crosswinds have on the size of safety margins under a variety of operational and environmental conditions; both with and without new flight deck technologies.

Method

Twenty-four pilot crews participated in this research: 12 airline crews and 12 business jet crews, who were deemed proficient in using a HUD. For normal operations, five levels of

Type of Guidance, three levels of RVR (300ft, 500ft, & 700ft), and two levels of Lighting conditions were examined (see Table 1). Wind speeds ranging between 3kt (calm) and 22kt[2] and varying directions were randomly assigned to scenarios. For abnormal operations, winds between 3kt (calm) and 15kt were applied. All tailwinds were limited to 10kt (Boeing 737- 800NG Airplane Flight Manual Limitation).

Table 1. Normal and Abnormal operational conditions

Operation Type	**Conditions**	**RVR**		
		300	**500**	**700**
Normal Operations	**Baseline 1**: HUD, No LOC[1] guidance, RCLM[2] only	Day/ Night	Day/ Night	Day/ Night
	Baseline 2: HUD, No LOC guidance, CLL[3,4]	Day/ Night	Day/ Night	Day/ Night
	Baseline 3: No HUD, CLL	Day/ Night	Day/ Night	Day/ Night
	Condition 1: HUD, LOC guidance, RCLM only	Day/ Night	Day/ Night	Day/ Night
	Condition 2: HUD, LOC guidance, No RCLM, No CLL	Day/ Night	Day/ Night	Day/ Night
Abnormal Operations	**Failure Condition 1:** Above V_1 Continue	Day/ Night	Day/ Night	Day/ Night
	Failure Condition 2: Below V_1 Reject	Day/ Night	Day/ Night	Day/ Night
	Failure Condition 3: Below V_{mcg} Reject	Day/ Night	Day/ Night	Day/ Night
	Failure Condition 4: LOC Fail	Day/ Night	Day/ Night	Day/ Night
	Failure Condition 5: LOC Bend	Day/ Night	Day/ Night	Day/ Night
	Failure Condition 6: Loss of HUD	Day/ Night	Day/ Night	Day/ Night

Crew workload was assessed using the "paper and pen" NASA-TLX [4]. The Pilot Flying (PF) and Pilot Monitoring (PM) each completed the NASA-TLX questionnaire immediately

[1] LOC = localizer
[2] RCLM = Runway Centerline Markings
[3] CLL = Center Line Lights
[4] All CLL conditions assume existing RCLM

following each takeoff scenario. Plotting the PF and PM as separate lower boundary layers is prudent and essential to the understanding of the overall nature of safety margin. More specifically, it affords a deeper insight to the changes of its shape and size as the crew experiences the changes in the operational and environmental conditions.

Our previous research [7] analyzed total weighted NASA-TLX scores, using an ANCOVA method, with two significant main factors: Type of guidance (five levels), RVR (three levels), and one non-significant main factor of lighting conditions (two levels). Crosswind component was considered as a covariate. In this study, the raw NASA-TLX subscale scores were examined separately, and utilized to construct 3-D safety margin profiles. The profiles show the changes that occur to the shape of the safety margin's lower boundaries under the different experimental conditions.

Due to violations of normality and homogeneity of variance, the crosswind coefficient for each subscale was evaluated individually using a Generalized Linear Mixed Model (GLMM). Instead of applying a transformation to the workload data, an identity link function to a normal distribution was applied. The crosswind component values were normalized across all experimental conditions; assessing at ~ 9kt for normal operations; and ~4kt for abnormal operations.

Using mesh graphs, the raw and the standardized NASA-TLX subscale scores for the PF and PM were plotted side-by-side. Crosswind component at ~ 9kt for normal operations and ~ 4kt for abnormal operations were applied with each crosswind component coefficient. In lieu of a linear interpolation method, the Modified Akima (MAKIMA) piecewise cubic Hermite interpolation method [1] in MATLAB® was utilized to generate the 3-D mesh graphs. For the purposes of 3-D visualization, this interpolation method allows for optimal data smoothing while at the same time preserving data validity.

Results

Crosswind component coefficients were determined for each NASA-TLX subscale by conducting GLMM analyses on the scores for normal and abnormal operations. The crosswind coefficients and associated p-values, for each NASA-TLX subscale for normal operations are shown in Table 2. The coefficient of crosswind covariate was positive and significant for each subscale. Table 3 shows the crosswind coefficients and associated p-values, for each NASA-TLX subscale for abnormal operations. The crosswind components coefficients were significant only for Temporal Demand and Frustration.

Table 1.
Crosswind Component Coefficient and p-value for each NASA-TLX Subscale (Normal Operations)

NASA-TLX Subscale	Crosswind Component Coefficient	*p*-Value
Mental Demand	0.459	$p < 0.0001$
Physical Demand	0.313	$p < 0.0001$
Temporal Demand	0.294	$p < 0.0001$
Performance	0.284	$p < 0.0001$
Effort	0.473	$p < 0.0001$
Frustration	0.362	$p < 0.0001$

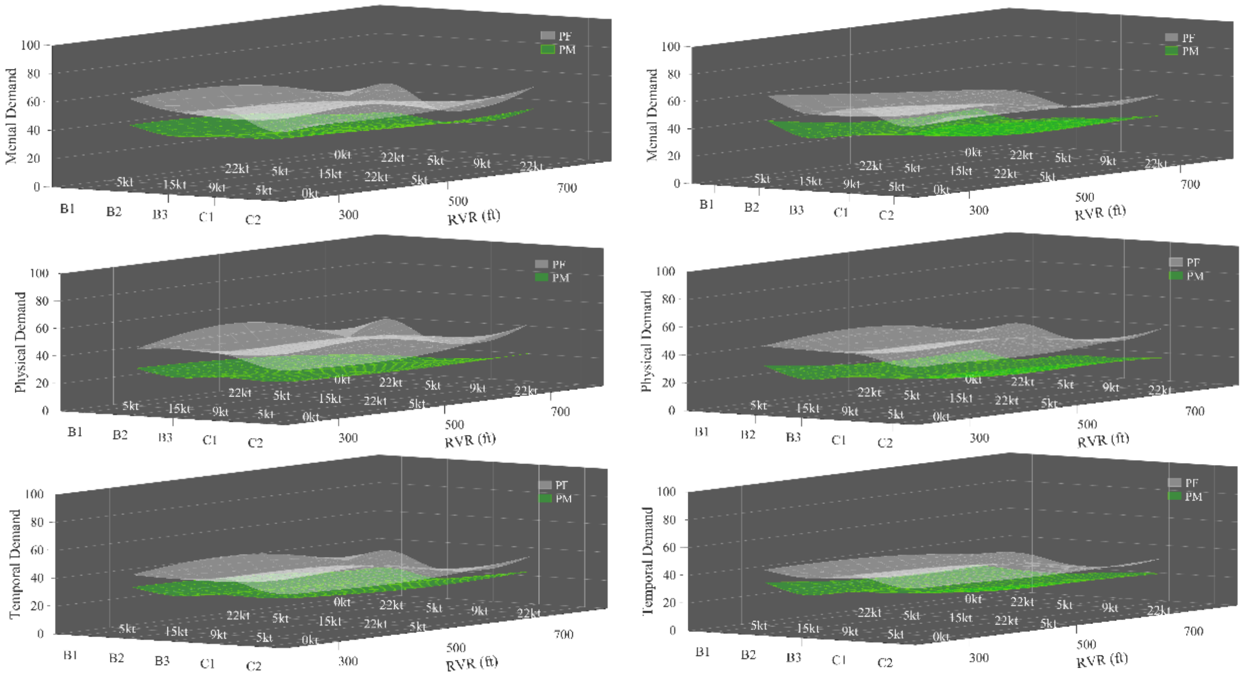
Table 2.
Crosswind Component Coefficient and p-value for each NASA-TLX Subscale (Abnormal Operations)

NASA-TLX Subscale	Crosswind Component Coefficient	*p*-Value
Mental Demand	0.279	$p = 0.170$
Physical Demand	0.138	$p = 0.476$
Temporal Demand	0.454	$p = 0.044$
Performance	0.236	$p = 0.141$
Effort	0.118	$p = 0.601$
Frustration	0.351	$p = 0.060$

A set of 3-D crew workload profiles for each NASA-TLX subscale across different Types of Guidance, RVR levels, and crosswinds, during normal operations at night are shown on Fig. 2 and Fig. 3. The workload profiles represent the lower boundary of the safety margin for a given set of flight conditions. The side-by-side profiles are grouped by NASA-TLX subscale, displaying the raw scores before parsing out crosswind effect on the left and the scores with the standardized crosswind effect on the right. Each figure represents a group of three subscales, one for the task demands subscales and one for the subscales related to the subjective experience of the person performing the task.

These profiles reveal the complex interplay between different factors affecting the lower safety margin boundary. The raw NASA–TLX subscale scores for PF and PM identify the sources of crew workload with a high level of granularity. For example, Fig. 2 and Fig. 3 clearly indicate that Mental Demand and Effort were the largest contributors of workload for both PF and PM in normal operations. The relatively higher “peaks” on the virtual “landscape” as shown on Fig. 2 (left) and Fig. 3 (left), portray the effect of crosswind component on the raw NASA-TLX scores in each set of conditions. In contrast, Fig. 2 (right) and Fig. 3 (right) depict the “landscape” when the effect of crosswind was parsed out.

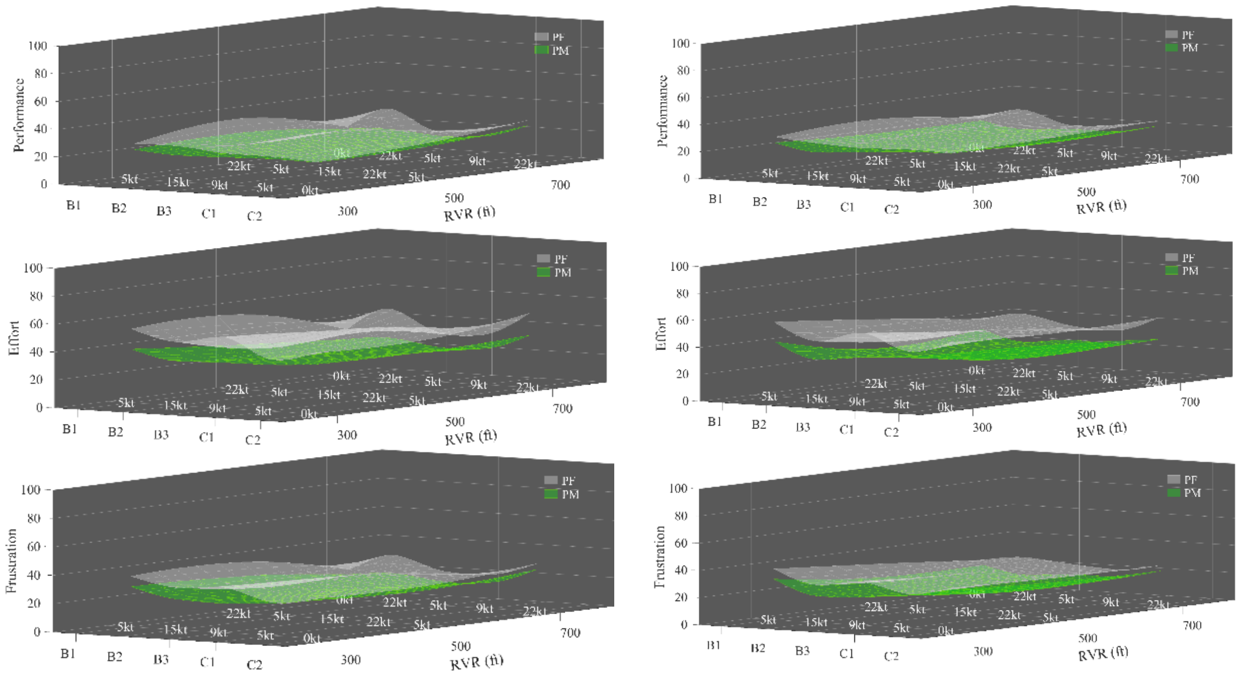
While the higher “peaks” on the raw score profiles could be interpreted as a decrease in safety margin under these conditions, none of them approached what would be considered unsafe levels of workload (e.g., scores in the upper most quartile on the NASA-TLX scale). We attribute these results to the high levels of information redundancy typical of the design of modern flight decks and the number of safeguards in place for multi-crew operations such as continuous training, practicing good CRM; and new technologies that enhance pilot performance and optimize workload. Nonetheless, having the "landscape" of crew workload visualized in 3-D space provides for a better understanding of the nature of the effect crosswinds (or any other factor of interest) may have on safety margin when interacting with conditions such as inclement weather during a critical phase of flight.



Legend:
B1 – Baseline 1: HUD, no LOC guidance, RCLM only
B2 - Baseline 2: HUD, No LOC guidance, CLL
B3 - Baseline 3: No HUD, CLL

C1 - Condition 1: HUD, LOC guidance, RCLM only
C2 – Condition 2: HUD with LOC guidance, no RCLM, no CLL

Fig. 1. NASA-TLX Mental Demand, Physical Demand, Temporal Demand (normal operations): raw scores (left) and standardized scores at ~9kt crosswind component (right)



Legend:

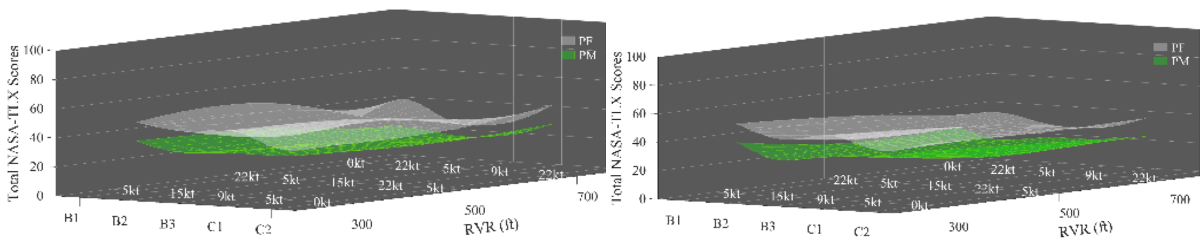
B1 – Baseline 1: HUD, no LOC guidance, RCLM only
B2 - Baseline 2: HUD, No LOC guidance, CLL
B3 - Baseline 3: No HUD, CLL
C1 - Condition 1: HUD, LOC guidance, RCLM only
C2 – Condition 2: HUD with LOC guidance, no RCLM, no CLL

Fig. 2. NASA-TLX Performance, Effort, Frustration (normal operations): raw scores (left) and standardized scores at ~9kt crosswind component (right)

The crosswind conditions included in the failure scenarios did not exceed the maximum authorized 15kt crosswinds for takeoff operations performed according to FAA Policy Order 8400.13F [3] and OpSpec CO78/079. The operational safeguards specified in these documents are designed to ensure safety margins are preserved and the maximum authorized crosswinds remain below the limits of both the aircraft and the pilot, especially when other risk factors such as inclement weather are present. Not surprisingly, crosswinds did not have a significant contribution to the level of crew workload except for Temporal Demand. One plausible explanation for these results could be that when a failure condition was present, the crew reprioritized crosswind as a factor. Their focus was on quickly resolving the effects of key factors jeopardizing safety of flight (e.g., engine failure) first, and then, on successfully recovering from the failure.

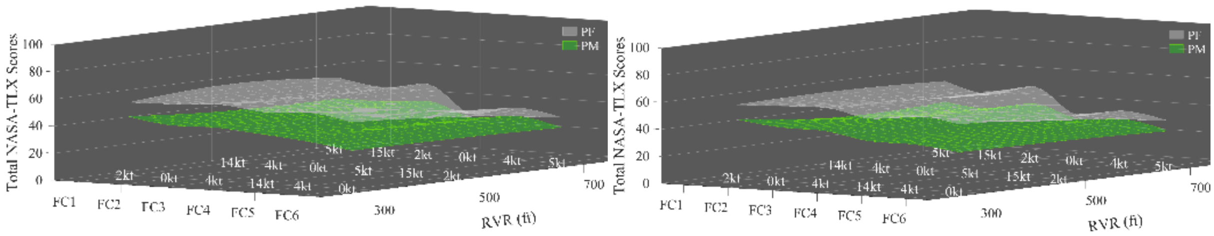
Discussion

In the original research effort [7], only the total weighted NASA-TLX scores were used for the analyses, and crosswinds were not a factor of primary interest. The overall crew workload levels reported by the crews across the baseline and experimental conditions were low to moderate for normal operations (Fig. 4). Workload did not exceed moderate levels on the NASA-TLX scale for abnormal operations (Fig. 5). the NASA-TLX scale for abnormal operations (Fig. 5).



Legend:
B1 – Baseline 1: HUD, no LOC guidance, RCLM only
B2 - Baseline 2: HUD, No LOC guidance, CLL
B3 - Baseline 3: No HUD, CLL
C1 - Condition 1: HUD, LOC guidance, RCLM only
C2 – Condition 2: HUD with LOC guidance, no RCLM, no CLL

Fig. 3. NASA-TLX Total Weighted (normal operations): raw scores (left) and standardized scores at ~9kt crosswind component (right)



Legend:
FC1 - Above V_1 Continue
FC2 - Below V_1 Reject
FC3 - Below V_{mcg} Reject
FC4 - LOC Fail
FC5 - LOC Bend
FC6 - Loss of HUD at V_R

Fig. 4. NASA-TLX Total Weighted (abnormal operations): raw scores (left) and standardized scores at ~4kt crosswind component (right)

The novel 3-D visualization proposed here utilized the scores on the six NASA-TLX subscales. The differential contribution of each source of workload to the crew workload profiles under the different experimental conditions was clearly identifiable when analyzed and plotted separately instead of using the weighted total NASA-TLX scores. This approach also allowed for a better insight to the effects all experimental factors had on the overall shape of safety margin's lower boundary.

In addition, during the data collection for the original research [7], we requested that pilot evaluators give their personal recommendation for the lowest RVR they considered as equally safe for using HUD localizer guidance symbology in lieu of centerline lights (CLL). Their feedback was remarkably consistent with the results and expanded beyond the topic of equivalence of level of safety to cover many other factors that could potentially influence it. Below are excerpts from the pilots' responses directly addressing crosswind as an important safety factor in that context.

"Successful outcomes were demonstrated at crosswinds of 22kt which is approaching the upper boundaries for a successful low visibility takeoff. I personally would suggest an operational limit of no greater than 15 knots crosswind component if utilizing an RVR value < 1800 for the Boeing 737 aircraft."

"Guidance cue was extremely helpful, mostly during crosswinds and engine failures." "I found it most useful making initial crosswinds corrections."

"I believe that the cue is a little too sensitive to very small rudder inputs in maintaining centerline in crosswinds."

"Guidance cue was extremely helpful, mostly during crosswinds and engine failures."

"The guidance que and ground roll reference cue appeared to diverge from runway centerline on crosswinds. Guidance cue seemed overly sensitive on initial takeoff roll. Training would be critical component of implementing this on 121 programs."

"Crosswinds affected the symbology on the HUD at rotation and takeoff by slewing rapidly and resulting in momentary confusion. After seeing the flight path vector slew to the side, the flight guidance cue eventually centered. It was only 1-2 seconds of uncertainty but because we rarely see very low visibility combined with strong crosswinds, the momentary confusion may cause over-controlling on the part of the PF."

One limitation of our research was the absence of wind gust conditions. Wind gusts are an important factor that could have a significant effect on the margin of safety [8]. However, creating wind gusts with a reasonable level of fidelity in a simulator presents a challenge. The ability to include wind gusts in a study conducted on a simulator may provide even more insightful lower boundary profiles with potentially more dramatic "peaks" and "valleys" representing the effects of wind gusts, in addition to the effects of sustained winds direction and magnitude, on the shape and size of safety margin.

Conclusion

A 3-D visualization of safety margin profiles, with and without the effect of crosswinds, has the potential to make the process of pilots setting up their individual crosswind limit more informed and for the following reasons:

- In a given set of operational and environmental conditions, specific sources of workload relevant to a piloting task in such conditions can be readily identified and considered;
- The "peaks" and "valleys" of the workload profiles for each potential source (e.g., Mental Demand, Temporal Demand) provide an insight to the potential changes to these profiles as conditions change;
- The safety margin profiles could provide a foundation for building robust mental models about how the effects of crosswind interact with the effects of other operational and environmental factors, as well as new technologies;
- Safety margin profiles may differ across different phases of flight, therefore understanding such differences, especially for takeoff and landing; could help pilots setting sensible crosswind limits.

From a broader perspective, the analyses conducted during this follow-on study, highlighted the notions that a) building an evocative shared mental model of the multi-faceted nature of safety margin, and b) having a clear understanding of the complex interplay between the many factors that affect its shape and size; are essential for the safety of flight.

While limited only to the takeoff phase of flight, these results demonstrated the potency of the NASA-TLX scale and the strong diagnostic value of its component subscales. The high reliability, sensitivity, and utility of these component ratings allowed for a narrow identification of the sources of workload in the specific set of conditions included in the research. However, a similar approach could be applied to other measures of pilot performance (e.g., flight technical error), as well.

Further research is need to expand the building of data-driven 3-D safety margin profiles across different phases of flight, operational and environmental conditions. Such profiles may be just the right tool to include in training materials and utilize in the cultivation of a culture of safety within the pilot community about something that is inherently subjective – perception of risk. Furthermore, as aircraft design involve multiple layers of safety margins intended to improve safety without unnecessarily limiting aircraft and human performance, designers could utilize such data-driven profiles to inform airframe and aircraft systems design, as well.

References

1. Akima, H.: A new method of interpolation and smooth curve fitting based on local procedures. Journal of the ACM (JACM) 17(4), 589–602 (1970).

2. Federal Aviation Administration, & United States.: Pilot's handbook of aeronautical knowledge. Aeronautical decision-making (pp. 2-1 – 2-32) Federal Aviation Administration, United States (2016).

3. Federal Aviation Administration.: Order 8400.13F. Department of Transportation, United States (2019).

4. Hart, S. G.: NASA task load index (TLX) Volume 1.0; Paper and pencil package. NASA Ames Research Center (1986).

5. Hart, S. G.: NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50(9), 904–908 (2006).

6. Hart, S. G., & Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in Psychology 52, 139–183 (1988).

7. Kratchounova, D., Humphreys, M., Miller, L., Mofle, T., Choi, I., & Nesmith, B. L.: Crew workload considerations in using HUD localizer takeoff guidance in lieu of currently required

infrastructure. In: International Conference on Human-Computer Interaction, pp. 507–521. (2020).

8. van Es, G. W. H.: Crosswind certification - How does it affect you?, (https://reports.nlr.nl/bitstream/handle/10921/343/TP-2006-324.pdf?sequence=1), last accessed 2021/2/4.