

COMPARISON OF HUMAN ICE DETECTION CAPABILITIES AND GROUND ICE DETECTION SYSTEM PERFORMANCE UNDER POST DEICING CONDITIONS

Scott M. Terrace & Kimberlea D. Bender
L-3 Communications – Titan Group

Edmundo A. Sierra, Jr.
FAA, Human Factors Research and Engineering Division

Isabelle Marcil
Transport Canada, Transportation Development Centre

John D’Avirro
APS Aviation Inc.

Edward Pugacz
FAA Office of Aviation Research, Flight Safety Branch, William J. Hughes Technical Center

Frank Eyre
Transport Canada, Transportation Development Centre

Currently, after deicing operations, the presence of residual ice on an aircraft’s wing is determined by a human deicer from a deicing ground crew via visual and tactile inspections. One method proposed to overcome some of the safety and physical concerns associated with human inspections is to use Ground Ice Detection Systems (GIDS). However, before regulatory authorities can consider GIDS for operational use, their performance had to be compared to human ice detection capabilities. In August 2005, the Federal Aviation Administration (FAA) William J. Hughes Technical Center’s (WJHTC) Simulation and Analysis Group conducted a study sponsored by the FAA Office of Aviation Research, Flight Safety Branch (WJHTC), and Transport Canada’s Transportation Development Centre to compare human ice detection performance using current visual and tactile techniques with the performance of two different GIDS under post deicing inspection scenarios. Nine male deicers from Globe Ground at Toronto Pearson Airport and Aero Mag 2000 Montreal performed post deicing inspections using three methods: the current method (Visual inspections and Tactile inspections), the GIDS1 method, and the GIDS2 method. Three separate post-deicing scenarios were presented each day for three days: a wing with 12 ice patches (High Contamination), three ice patches (Low Contamination), and a clean wing (No Contamination). Accuracy data, false detection data, and time to complete an inspection were collected and analyzed for each condition. The results from the study consistently indicated that overall GIDS1 was superior to human visual and tactile inspections and GIDS2 inspections in terms of accuracy, false detections, and stability in performance. Participants using GIDS1 were able to detect all patch sizes and thicknesses with the greatest accuracy while the other methods’ accuracy improved as a function of patch size and thickness. In addition, inspections completed by the GIDS1 manufacturer throughout the study suggest that, with time and experience, performance could further improve.

INTRODUCTION

Currently, after deicing operations, the presence of residual ice on an aircraft’s wing is determined by a human deicer from a deicing ground crew. The presence of ice on a wing is determined visually under most circumstances. Canadian and US regulations require tactile inspections following deicing for certain types of aircraft. Some problems have been identified with tactile inspections. Tactile inspections expose extremities to cold surfaces, require close proximity to an aircraft (at times with engines running), are slow, and can be limited by the deicer’s reach.

The Federal Aviation Administration (FAA) and Transport Canada (TC) are exploring the potential to supplement or replace human visual and tactile inspections with remote Ground Ice Detection Systems (GIDS). A GIDS Regulatory Approval Working Group (RAWG), under the

auspices of the SAE Committee G-12 Ice Detection Subcommittee, was formed to explore this possibility.

An initial threshold experiment was completed in March of 2005 (Sierra, Bender, Marcil, D’Avirro, Pugacz, & Eyre, in press). The threshold study attempted to quantify human visual and tactile ice detection capabilities to serve as a measure against which GIDS can be evaluated. Results from the threshold study were used to help determine the test parameters for the comparison study. If visual and tactile inspections for the presence of ice on a wing are to be replaced with GIDS, these systems must be as good as, if not better, at detecting the presence of residual ice than human visual or tactile capabilities.

In this study, two different GIDS were used for evaluation: the Ice Camera by MacDonald Dettwiler and Associates (MDA) and the Goodrich IceHawk[®] by Goodrich Aerospace. This paper is a condensed version of the full

length comprehensive study (Bender, Sierra, Terrace, Marcil, D'Avirro, Pugacz, & Eyre, 2006). The full length report discusses the study in more detail.

Participants

This study employed nine male deicing crew participants ranging from 25 to 53 years of age. Deicers ranged from 1 to 24 years of experience. The different experience levels of the participants had no impact on the results.

Far visual acuity was determined using a 20 foot Snellen Eye Chart. Three Deicers' corrected vision was worse than 20/20 (measured at 20/25, 20/40, and 20/25 respectively). All had normal color vision as determined by the Quick Six Color Vision Test.

Tactile discrimination ability was determined with the Grit Ordering Test (GOT), which was developed specifically for this series of experiments (Sierra et al.). For the GOT, deicers were asked to indicate the order of roughness of three sandpaper strips (400, 600, and 1500 grit), from least to most rough. One participant deicer failed the task. The data from this participant was within the range of the other participants and did not affect the data.

Materials and Apparatus

We conducted this study in the large PMG Test and Research Centre climatic chamber in Blainville, Quebec, Canada. The chamber dimensions were 54 feet long x 21.5 feet wide x 13 feet high. The temperature in the chamber was -5°C ($\pm 0.5^{\circ}$), humidity was 90% ($\pm 5\%$). No precipitation was used.

We attempted to replicate dusk/nighttime conditions. After consulting with subject matter experts (SMEs), two diffused, 150 watt high pressure sodium bulbs with approximately 14,000 mean lumens were deemed appropriate to light the chamber.

Personnel from APS Aviation Inc. formed ice patches in different location on an aluminum Lockheed JetStar wing (as depicted in Figure 1). A layer of diluted Type I deicing fluid was applied over the entire wing, including the ice patches, in order to simulate post deicing ice conditions. Ice patches were smooth with little to no edge. The ice patches varied with respect to size and thickness. We chose patch sizes and thicknesses that had low and moderate chances of being detected by the deicers. Subject matter experts (SMEs) estimated that 8 inch and 16 inch diameter ice patches were adequate for the low range and moderate range, respectively. Furthermore, the 8 inch diameter ice patch was used because it correlated to the area frozen contamination must cover to constitute deicing fluid failure during deicing fluid holdover time testing. Ice thickness varied among two ranges, 0.3 - 0.5 mm (Low detectability) and 0.6 - 0.8mm (Med. detectability). (For details about ice sample preparation, see Narlis, 2005.)



Figure 1. APS staff worker forming a 16 inch patch on the JetStar aluminum wing.

Two GIDS were used for this study. One of the GIDS, The Ice Camera, made by MDA, utilizes a multispectral infrared camera that detects both ice and water. The Ice Camera employs a reflectance spectroscopy technique to detect ice 0.5 mm or thicker (Gregoris, Yu, & Teti, 2004). The Ice Camera is able to remotely detect ice and display the images to enable the deicer to determine if the wing is still contaminated. If ice is detected, the image will be displayed as a red overlay on a grayscale image.

The other GIDS, the Goodrich IceHawk[®], developed by Goodrich Aerospace, uses a collimated laser light source to illuminate a small spot on the surface to be scanned with linearly polarized light. If light is reflected from this spot and is still linearly polarized, the surface is categorized as clean. However, if the reflected light is de-polarized in a certain way, the surface is considered contaminated with ice, frost, or snow. A series of these spot images are taken using a raster mirror to provide a camera field of view of $30^{\circ}\text{X}20^{\circ}$ using 60,000 spots or pixels for a 300 X 200 pixel image. When there is no ice present, the scan has been programmed to show a green scale image of the area examined. When ice, frost, or snow is present, the image will display red in those areas where the frozen contamination is present.

Procedure

Three test scenarios were used for this study. A within-subjects design was used; therefore each participant was tested using all ice detection methods for each test scenario, a High Contamination condition consisting of 12 ice patches placed on the wing, a Low Contamination condition with three ice patches on the wing, and a No contamination condition without any ice patches on the wing (see Table 1). The No Contamination condition was used to collect false detection data. For each test scenario and methods of inspection, the locations of the ice patches were randomized so that each participant could not memorize where the patches were in a previous session to get better results. Treatment condition order, subject order, and manufacturer order were counterbalanced for order effects. Table 1 shows the counterbalanced assignments for the nine participants.

Table 1. Deicer Assignments

Ice Detection System			
Day	Control	GIDS 1	GIDS 2
Training	1-9	1-9	1-9
1	1-3	7-9	4-6
2	7-9	4-6	1-3
3	4-6	1-3	7-9

Deicers were sorted into groups of three and assigned to a condition each day. They rotated through each condition on successive days in order to experience each one: for example, as depicted in Table 1, deicers 1 through 3 were in the control condition (visual and tactile inspection) on Day 1, in the GIDS2 condition on Day 2, and the GIDS1 condition on Day 3.

The day prior to the start of the study was dedicated to training and briefing the deicers. The GIDS manufacturers provided hands-on training to all deicers. Neither the trainers nor the Test Administrators revealed the manufacturers of the equipment. GIDS systems were referred to simply as GIDS1 and GIDS2 throughout the study. They are also referred to as such in the results section of this document.

Deicers were instructed to conduct inspections as accurately as possible and to be mindful of the time that they used to conduct the inspection. Accuracy data (number of patches correctly detected), and false detection data (number of patches identified that were not present) were collected and analyzed for each condition. Inspection times were also recorded. However, inspection times are not reported in this paper because the designs of the GIDS user interfaces did not have a speed requirement for this study.

Due to limited space in the chamber, the GIDS cameras were mounted in a fixed location throughout the test, limiting the systems to one distance and angle. Output for each GIDS manufacturer was transmitted to their own remote station. GIDS stations were arranged so that inspections were performed independently and deicers could not see each other's GIDS output. Figure 2 shows the placement of both GIDS and the scissor lift overlooking the wing in the chamber.



Figure 2. Placement of the GIDS systems, scissor lift, and wing in the chamber. Image was taken during preparation for the experiment.

To avoid giving deicers an advantage over the GIDS, inspection location was controlled by having the deicers performed their visual inspection from a point collocated with the GIDS. Visual inspections were performed from a scissor lift five feet from the wing allowing the deicer limited lateral movement. Deicers were free to move the scissor lift vertically if they chose. In addition, the deicers were free to do their inspections from the angles they normally use during post deicing inspections (i.e. they could crouch or swivel). For the tactile inspection, the deicers walked around the wing to perform the inspection. Deicers were allowed to visually scan the wing as they conducted their tactile trials and vary their viewing distance and angle. No tools were supplied to assist the operators in the inspections (e.g. ladders, stools, flashlights, or tactile wands).

Tactile inspections were performed as deicers would in the field. They were instructed to use open hand only, without scratching, to preserve the test samples. Deicers were allowed to begin the inspections at whatever part on the wing they begin with in the field. All tactile inspections were performed with gloves on. Deicers were allowed to visually inspect the wing as they performed the inspection because they are able to do so in the field. They could reach as far into the wing as they chose, but no stools or equipment were provided to help them extend their reach.

Results

After verifying that the assumptions for the data were met, a one-way repeated measure Analysis of Variance (ANOVA) was conducted on the high contamination data to determine if method of inspection had a significant effect on the correct number of ice patches detected. An alpha level of .05 was used for all statistical tests.

Results of the repeated measures ANOVA indicated a significant difference between the methods of inspection for the number of correct ice patches found, $F(3, 24) = 23.59, p < .05$. Figure 3 summarizes the means of each method of inspection. Since there was a significant effect of inspection method found, pairwise comparisons were analyzed using the Least Significant Difference (LSD) method.

Results for the significant pairwise comparisons revealed that participants using GIDS1 found significantly more patches ($M = 10.33, SD = 1.73$) than GIDS2 ($M = 5.11, SD = 1.76$), $p < .05$. Participants using the GIDS1 method found significantly more patches ($M = 10.33, SD = 1.73$) than the Visual ($M = 4.44, SD = 1.67$) and Tactile ($M = 7.11, SD = 1.62$) methods of inspection, $p < .05$. Participants using the Visual method found significantly fewer patches ($M = 4.44, SD = 1.67$) than the Tactile method ($M = 7.11, SD = 1.62$). $p < .05$. See Figure 3 for summary data.

High Contamination

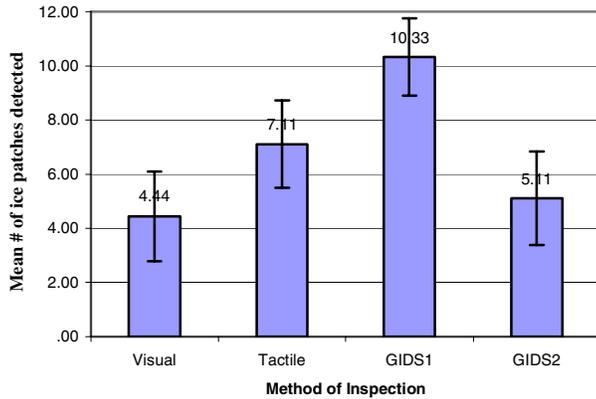


Figure 3. Mean number of ice patches detected for each method of inspection during the High Contamination condition.

For total patches detected in the Low Contamination condition, a non-parametric Friedman's analysis of variance by ranks test was used since the data summarized in Figure 4 did not meet the assumptions of normality to conduct parametric tests.

Low Contamination

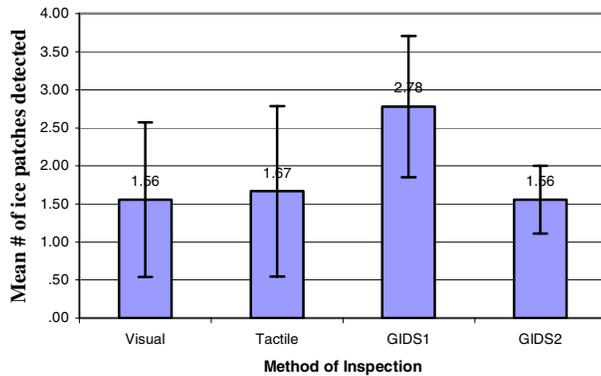


Figure 4. Mean number of ice patches detected for each method of inspection during the Low Contamination condition.

The Friedman's analysis of variance also resulted in a significant overall difference for the method of inspection variable $F(2) = 11.33, p < .05$. Table 2 shows the mean ranks for each of the four methods of inspection for this analysis. Order of mean rank from highest to lowest was GIDS1 (3.61), Tactile (2.28), both Visual and GIDS2 (2.06). To identify where the group differences were located, a Friedman's test for multiple comparisons was used. No significant pairwise comparisons were found, $p > .05$.

Table 2. Mean ranks for each method of inspection.

Method of Inspection	Mean Rank of patches found
Visual	2.06
Tactile	2.28
GIDS1	3.61
GIDS2	2.06

Participants using GIDS1 were able to detect all patch sizes and thicknesses with the greatest accuracy while the other methods' accuracy improved as a function of patch size and thickness (see Figure 5).

High Contamination

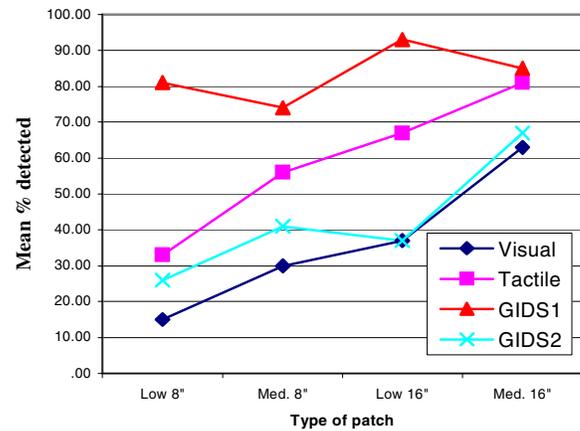


Figure 5. Mean percent correct detection of each type of ice patch for all methods of inspection.

The data for detecting an ice patch when there was not one present (false detection) was analyzed for the High and Low contamination tests. A false detection was counted when there was no ice present on a specific location on a wing and the deicer indicated there was ice present.

For the High Contamination test, GIDS1 resulted in the fewest false detections (one), followed by the Visual and Tactile method (three and three, respectively), and lastly GIDS2 (four).

For the Low Contamination test, the Visual condition resulted in the most false identifications (11), followed by the Tactile condition (four), GIDS2 (one), and GIDS1 (none).

In a separate additional test, we compared GIDS manufacturer results to deicer results in order to ascertain whether experience and training with the system might influence results. The GIDS1 manufacturer found all 12 patches on the wing for all three high contamination runs although not all of the deicers did (see Figure 6). This suggests that training and experience may play a role in using the equipment. The GIDS2 manufacturer representatives were unable to locate all of the patches for the high contamination runs suggesting that the equipment itself was not detecting the patches and training is not necessarily the only factor influencing the ability to detect ice with this system.

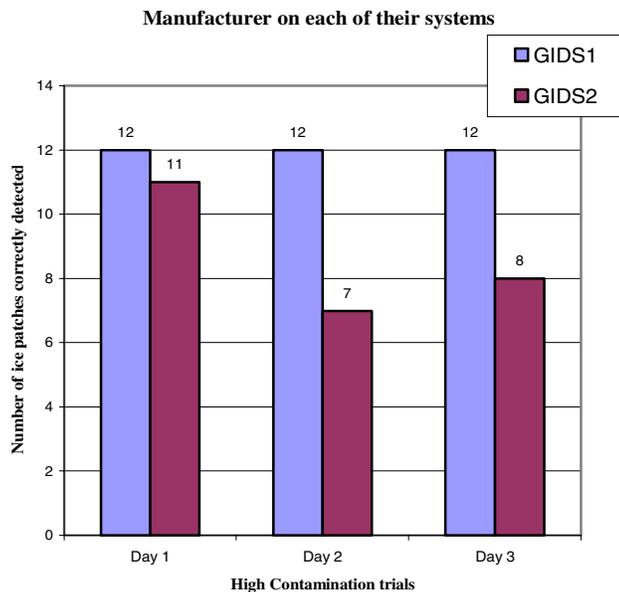


Figure 6. Number of patches correctly detected by the GIDS manufacturers for the High Contamination Trials.

Discussion

Accuracy evaluations were conducted through two different tests: the High Contamination and Low contamination tests. Analyses from both tests were consistent, in that GIDS1 detected more patches than the other methods for both tests. GIDS1 was also superior to the other methods of inspection in terms of stability in performance, as it was able to detect all patch sizes and thicknesses with the most accuracy. Visual, Tactile, and GIDS2 methods had lower accuracy rates for the thin patches (.3-.5 mm) in particular. Performance did appear to improve as patch size and thickness increased for the Visual, Tactile, and GIDS2 conditions. In addition, GIDS manufacturers conducted inspections prior to deicer runs for the purpose of collecting supplemental data. The resulting data from these inspections indicate that with time and experience, GIDS1 performance could further improve.

Data analysis also indicated that the Tactile inspections found more patches than the Visual inspections or GIDS2 inspections for the high and low contamination tests, although the difference was not always statistically significant. The Tactile test simulated in this study was conducted on a wing that was relatively low to the ground. A higher wing may have further limited the reach of the operator, introducing the possibility that detection may have decreased. Visual inspections were similar to GIDS2 inspections in terms of accuracy. Visual and GIDS2 tests results were comparable for the low contamination test but the GIDS2 systems performed slightly better for the high contamination test. However, it should be noted that visual inspection performance in our study may be lesser than an actual visual inspection since they were conducted from a fixed location.

Despite the fact that statistical significance was not always reached for the accuracy data, it is important to note

that the consistent overall differences, as well as significant differences, may constitute operational significance. The reliability of the data across tests suggests that the potential for operational improvements may exist with use of the GIDS1 system.

The results from the study indicate that a technology currently exists that is as good as or better than the current human detection system. This study was limited in that the GIDS were not field ready and could not be evaluated properly for detection speed. Once these systems are in a field ready state, they should be compared again for detection speed.

According to the results from this study, there is strong evidence that the GIDS1 system should be further evaluated to ensure that detection capability is stable across a variety of conditions. Furthermore, the GIDS1 system should undergo an interface and design evaluation with SMEs providing input regarding system design prior to operational use in the field.

Study results indicate that GIDS2 should not be considered for field implementation in its current configuration. Technical improvements and further testing would have to be conducted before a recommendation for further consideration of this device in a field setting could be made.

Test results suggest that deicers may not meet the current standard of a 100% clean wing, 100% of the time (although we do recognize that the Visual inspections procedures used in this study were not necessarily as comprehensive as they might be in the field). The Visual inspection procedures used in this study did not account for movement, illumination, or other tools available to deicers in the field. However, conditions in the field will sometimes be much more difficult than the conditions used in the study. Further testing with realistic and comprehensive visual and tactile inspections could be performed if this type of field testing is to be pursued.

REFERENCES

- Castellani, J.W., O'Brien, C., Baker-Fulko, C., Sawka, M. N., & Young, A. J. (2001). *Sustaining health and performance in cold weather operations* (Technical Note No. TN/02-2). Natick, MA: US Army Research Institute of Environmental Medicine.
- Gregoris, D., Yu, S., and Teti, F. (May, 2004). Multispectral imaging of ice. *IEEE CCECE Conference, 2004*. Niagara, Ontario, Canada.
- Narlis, C. (2005). *Development of Ice Samples for Comparison of Human and Sensor Capability to Detect Ice on Aircraft* (TP 14450E). Montreal, Canada: Transport Canada Development Centre.
- Bender, K., Sierra, E. A., Jr., Terrace, S., Marcil, I., D'Avirro, J., Pugacz, E., & Eyre, F. (2006). *Comparison of Human Ice Detection Capabilities and Ground Ice Detection System Performance Under Post Deicing Conditions*. (Technical Report DOT/FAA/TC-06/20) Atlantic City International Airport, New Jersey: US Department of Transportation, Federal Aviation Administration.
- Sierra, E. A., Jr., Bender, K., Marcil, I., D'Avirro, J., Pugacz, E., & Eyre, F. (2006). *Human Visual and Tactile Ice Detection Capabilities under Aircraft Post Deicing Conditions*. Atlantic City International Airport, New Jersey: US Department of Transportation, Federal Aviation Administration.