**Robust Panel Set Definition for RS-25 FPI POD Demonstration Testing**

**Mindy Hotchkiss1**

1Aerojet Rocketdyne, an L3Harris Technologies Company

17900 Beeline Highway

Jupiter, FL 33458 USA

(561) 882-5331; mindy.hotchkiss@l3harris.com

ABSTRACT

The Aerojet Rocketdyne RS-25 program went through a recent effort (2019-2021) to replace the panel set used for Probability of Detection (POD) demonstration testing for Fluorescent Penetrant Inspection (FPI), due to the deteriorating condition of the existing NASA-owned specimen standard, at that time already in use for over 40 years and which had gone through multiple refurbishments. This legacy set was being further reduced in size in terms of the number of flaws available to inspection due to flaw growth (porosity). Additional investment into activities needed for further life-extension was therefore deemed inadvisable, so a new panel set had to be acquired.

The panel set acquisition process involved a number of practical decisions about panel and flaw characteristics, with many different considerations taken into account, e.g., what kind of flaws, how many flaws were needed, and what sizes. The focus of this discussion is flaw sizing and allocation, specifically how the target flaw distribution provided to the vendor was obtained. One key constraint was that, due to production uncertainty, the vendor would not necessarily be able to produce an exact flaw size match for all flaws requested, but would need to target intervals. As a result, panel set design planning needed to account for and accommodate this limitation, not in terms of the actual resolution desired.

A number of distributional options were then compared using this interval basis, on logarithmic scale, for overall utility and robustness to meet potential future needs, given the long-lived nature of the set. Existing guidance recommended that flaws be distributed uniformly on log scale for Level 3 and 4 sensitivity penetrants; however, once modeled this was found to have undesirable qualities that were suboptimal and potentially inefficient. A statistical Beta distribution was used to compare different distributions relative to the Uniform baseline, a subcase of the Beta, and which vary in both skew and kurtosis. A calculator tool provided quick visualizations and facilitated the decision-making process by enabling optimization of the target allocation to meet desired objectives.

Once the panel set had been manufactured and was available on-site in late 2020, the Materials and Process Engineering group conducted a detailed review on all panels to verify the dimensions of the identified flaws. The panels were processed multiple times to validate inspection capability. A surplus of flaws was obtained to provide margin against future deterioration and natural wear and tear on the panel set. Panels were also divided into overlapped sets to diminish overuse concerns, but with common panels to provide a mechanism for future panel and flaw health monitoring. A database was also designed to serve as a repository for the data, facilitate decision-making, and be a resource for the future.

**Keywords:** RS-25, Probability of Detection, POD, Fluorescent Penetrant Inspection, FPI, panel set design, flaw distribution, flaw allocation, uniform distribution, beta distribution

INTRODUCTION AND BACKGROUND

The National Aeronautics and Space Administration (NASA) has been developing the Space Launch System (SLS) heavy-lift launch vehicle in support of the Artemis lunar landing and exploration program, intended to establish a sustainable human presence on the moon. Fifteen retired Space Shuttle Main Engines (SSMEs), known as RS-25s, were repurposed with new design modernized engine controllers and a 16th engine was assembled from heritage Space Shuttle assets by Aerojet Rocketdyne (AR) in support of this effort [1, 2]. NASA Technical Standards establish requirements for fracture control of hardware associated with manned spaceflight, in order to mitigate the risk of catastrophic failure due to material flaws [3]. These standards outline specific processes and requirements for the implementation of various Nondestructive Evaluation (NDE) assessments, including Fluorescent Penetrant Inspection (FPI), the focus of this work [4].

For FPI, inspector capability is established quantitatively through administration of a Probability of Detection (POD) demonstration test, in which a set of panels with known flaws of different sizes is inspected. The resulting dataset of hit/miss outcomes is analyzed to estimate a POD score for that inspector using binary logistic regression, a standard statistical method [5]. Binary logistic regression uses a link function, such as logit or probit, to transform the binary outcomes to a probability scale, then obtains best-fit estimates of model parameters from the transformed data relative to the independent variable [6], which for hit/miss analysis is crack length. An inspector’s POD score, or , is the estimated crack length for which they are predicted to have a 90% probability of detection at 95% confidence. A unique score must be obtained for each penetrant-method combination [5].

The RS-25 program at AR completed a recent effort (2019-2021) to replace the panel set used by its FPI inspectors for their POD demonstration testing for identification of mechanical fatigue flaws. The panel set is needed by RS-25 NDE to qualify inspectors for Level 3 (High) and 4 (Ultra-High) penetrant sensitivities, each with Line of Sight (LoS) and Borescope (BS) methods. Demonstration testing takes place in-house at multiple sites, as well as numerous vendor facilities. All inspectors must re-qualify every 3 years, for each penetrant-method combination. Hardware is only accepted if inspected by a certified inspector who has achieved a POD score below the designated critical initial flaw size (CIFS) for that component, calculated analytically per fracture mechanics.

The legacy specimen standard had been in use for over 40 years and had undergone multiple refurbishments to extend its useful life. However, the condition of the fatigue flaws on the panel set was further deteriorating and additional investments in life extension activities were not considered value-added or even necessarily viable. The distribution of flaws on the panel set had also been a matter of concern due to difficulties achieving convergence in model fits for the logistic regression, so this was considered an opportunity to re-baseline the process to better position the program for the future. This discussion focuses on the statistical considerations behind how the flaw allocation for the new panel set was defined, as a critical aspect of the panel acquisition process.

PRELIMINARY PANEL SET DEFINITION

Initial design and development of panel set requirements had two phases: (1) programmatic practical decisions that were needed up front (e.g., vendor, flaw type, panel dimensions, and material), and (2) refining the request, specifically related to the total flaw count and how the flaws were distributed, or allocated, on the panels.

Programmatic Considerations

The original manufacturer of the legacy panel set was no longer available, so a new panel set vendor was necessary. Obtaining the legacy panel materials, Inconel 718 with AMS5593 plate, was not an issue. However, the legacy panel set had been produced using mechanical fatigue, but only thermally-generated fatigue cracks were presently available. The team considered this a disadvantage since thermally produced cracks were expected to be somewhat more easily visible to the inspectors, so anticipated POD scores could, in theory, improve slightly due to this inherent system bias in the tool. The legacy panels themselves were larger than the standard dimensions now available, so the team considered a custom order to replace the existing panels exactly (meaning fewer larger panels), as well as the new standard dimensions (more smaller panels). Acquiring the smaller panels provided some human factors advantages: easier transportability, less risk of loss or damage to the set if individual panels had issues, and limiting the ability of the inspectors to mentally retain information on flaw locations. The final full panel set received consisted of 52 smaller two-sided panels, which also includes blanks.

**Flaw Count**

A priority of the team was to provide sufficient flaws to provide margin against future losses, to mitigate the risks associated with losing flaws to the analysis that had been encountered with the legacy set, given the long life-span the specimens are expected to be in use. The team also wanted to narrow the uncertainty bands associated with the estimation process, so intended to use more than the recommended minimum 60 points. The target set requested from the vendor was therefore for a minimum of 80 flaws, to account for attrition over future years. The final total number of flaws obtained was 87, which provided the ample margin over the target minimum count desired.

**Flaw Allocation**

Selection of the flaw distribution for the panels was of critical importance to this project. More small flaws were needed, but how to allocate them presented challenges. The legacy distribution had a concentration of values (readily apparent on log scale) and insufficient small flaws to be robust, as well as several unnecessarily large flaws, which resulted in an unsatisfactory spread. The team’s goal to reduce the flaw max and flatten the distribution would enable the new panel set to better provide visibility into the threshold range of the POD curve. Inclusion of high points well beyond the range of interest was considered inefficient, so decreasing the max to a more reasonable crack length would allow all testing to be more value-added. The team was also concerned about the different flaw type being easier to detect, so needed to allow for additional smaller values since the magnitude of this effect was unknown. As a further complication, the vendor requested that the team provide the target allocation in intervals, which they would attempt to approximate, due to uncertainties in the flaw manufacturing process. As a result of this limitation, the panel set could not be fully optimized with respect to specific quantitative targets, but had to be defined in a way that would be robust to various stakeholder needs.

FLAW ALLOCATION METHODOLOGY

Available guidance suggested flaws be distributed uniformly on either log or Cartesian scale [5]. AR RS-25 standard practice has been to use log scale in POD analyses, to account for the measurement being bounded at zero (since crack sizes must be positive). No guidance was available on choice of the max, the upper bound, but inclusion of high points beyond the range of interest is not needed to mathematically anchor the model in any way. The team was uncomfortable with the recommendation of cracks being uniformly distributed, on either scale, and preferred to allow for some tapering on both sides of the flaw size range. The statistical Beta distribution [7, 8] was used for comparisons, since it has the flexibility to accommodate a range of shapes between two bounds, and includes the baseline Uniform distribution as a special case, which has constant probability between two bounds.

**Beta Distribution**

A generalized version of the Beta probability distribution function (PDF) is shown in Equation 1, on Cartesian scale.

**(Eq. 1)**

where

= shape parameter (),

= scale parameter (),

,

, and

, for z real >0 (Gamma function).

The Beta distribution shape between two bounds is defined by two positive-valued parameters and . With , the Beta is equivalent to a Uniform, so was a natural mechanism for comparison of numerous options. Both Uniform and Beta require the selection of a max crack size, but the Beta allows tapering by changing the and parameters. The model was applied on log scale.

**Flaw Allocation Visualization Tool**

A calculator tool was constructed to automatically allocate target Beta distributions in intervals across the specified range on log scale, provide a visual comparison to the legacy flaw allocation, as well as to the baseline Uniform. Results were color coded for easy visibility, so that subtle differences were more easily distinguishable. The distribution could be adjusted by modifying flaw max size and the number of desired intervals.

Flaw Size Distribution Limits

A panel set needs to have sufficient small flaws to anchor the distribution at the low end. Flaws at or below 0.010 in are found rarely, even by the best inspectors. The minimum flaw size in the initial panel set requirements was 0.005 in, the smallest flaw that could be requested from the vendor. The smallest flaw in the legacy set was 0.008 in.

The flaw distribution is defined in large part by its upper bound, so identifying an appropriate value was a critical input. The maximum flaw size in the legacy panel set was 0.417 in, which was considered unnecessary, given the target threshold range of interest was nominally at 0.050 in. A number of options were considered and sensitivities assessed. The max target flaw size considered in defining the initial panel set requirements was targeted to be 0.150 in, 3X the target nominal threshold range of 0.050 in, a notable decrease from the legacy max value. This 3X factor was considered sufficiently wide enough to capture the threshold range for the expected range of inspector capabilities, especially given the tighter bounds available with the higher flaw count. While the panel set redesign needed to be able to quantify performance for inspectors across all performance levels, a key focus was on improving the tool’s capacity to assess higher performers. Using a lower max provides greater visibility into that of the better inspectors, since it concentrates the flaws into a narrower zone.

Linear Flaw Density

A simple metric was constructed to characterize overall panel set attributes, independent of inspector performance. Linear flaw density, as defined in Equation 2, reflects average flaw levels over the linear range available.

**(Eq. 2)**

where

= linear density of flaws per unit length,

= total number of flaws on the panel set,

= linear range from maximum to minimum flaw size values, and

= minimum flaw size in the panel set, and

= maximum flaw size in the panel set.

The legacy panel set had experienced decreasing flaw density, as its flaw count had dropped from 64 to 59; its baseline value of 156 had gone to 144 flaws/in. A variety of options were considered. The target allocation of 80 flaws with max flaw size approximately equal to 0.150 in would have improved flaw density to ~550 flaws/in (90/0.145), assuming all were used, a substantial improvement over the legacy set. Flaw density should be optimized depending on specific NDE needs; higher flaw density is not a simple indication of quality.

QUALIFICATION AND FINAL PANEL SET DEFINITION

The AR Materials and Process Engineering (MPE) Lab was tasked to identify and characterize flaws found on the new panels. Optical microscopy was used to view and capture flaw images, with measurements obtained using the Keyence software. Documentation provided by the vendor was used to create masks for each panel using approximate locations and length. Differences were noted in measured values and orientation for individual flaws, but no consistent bias was present (paired t-test, p-value = 0.222). AR MPE measurements have been used for all subsequent work. The final distribution of flaws obtained contained 87 flaws, more than the target minimum count of 80 requested, with more smaller flaws than requested (Log-rank test, p-value = 0.006).

After receipt of the full panel set of 87 flaws on 52 panels, the team considered alternatives for use of the surplus samples. Two statistically equivalent (Log-rank test, p-value = 0.924) overlapping subsets were developed, each consisting of 72 flaws on 40 panels. Approximately 75% of the set is common to both subsets, with the remaining 25% assigned to one set only. Time expended on conducting the test is driven by the number of panels, but the width of the confidence bounds is driven by a combination of inspector capability and the number of sample flaws in the demonstration test population. Given the same inspector capability, having more flaws produces a tighter confidence limit. Development of the smaller panel subsets also enabled further modifications to the flaw distribution. The final allocation was flatter and compensated for the bias toward smaller values in the full set, as shown in Table 1.

Table 1: Panel set comparison shows increase in flaw density as flaw distributions were reallocated.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Panel Set | Flaw Count | Minimum | Median | Maximum | Flaw Density (Count / Range) |
| Legacy | 64 🡪 59 | 0.008 in | 0.069 🡪 0.067 in | 0.417 in | 156 🡪 144 |
| Full Set | 87 | 0.006 in | 0.036 in | 0.132 in | 690 |
| Subsets | 72 | 0.010 in | 0.048 in | 0.132 in | 590 |

IMPLEMENTATION AND DEPLOYMENT

The new panel subsets were deployed in June 2021 and as of February 2024 have been used for nearly 50 demonstration tests to date, both in-house and at multiple vendors. Neither subset has experienced any convergence issues in obtaining a valid fit to the logistic regression. Only a few inspectors have been tested on both the legacy set and the new subsets to date, but scores are internally consistent without making adjustments due to experience and do not suggest that any notable bias is present due to the change in the flaw mode of generation. A longer-term objective of this project is to use the new data collected for panel set health monitoring, in order to identify flaws that may be change over time, as has been seen in other legacy panel sets. A database is in development to serve both as a data repository and facilitate future decision-making.

SUMMARY

The RS-25 program relies on NDE processes for acceptance of flight-critical hardware. Getting the new FPI POD panel set defined and qualified was crucially important to qualifying both AR and supplier FPI inspectors and came at a critical time for the program, since continued use of the legacy panel set was no longer viable. Replacing the legacy set with a more robust version positioned the program to better support NDE work for decades to come. Using the Beta distribution for characterizing the flaw distribution provides a more flexible mathematical framework for investigating subtle differences between panel set options. Likewise, the flaw density metric provides a simple and intuitive measure for characterization of overall panel set quality that is independent of inspector capability.

ACKNOWLEDGEMENTS

This project was the work of a team led by NDE specialists Fred “Kon” Haake (retired) and Jeffrey Ragonese, QA, with support from David Erling and Frederick Batignani, NDE (retired); Jesse Harris, MPE Lab; and Justin Zwolski, Statistics. Also thanks to Fredrick M. Whitman, RS-25 program manager, whose support was instrumental in working through this process.

REFERENCES

1. National Aeronautics and Space Administration, 2024, “Space Launch System RS-25 Core Stage Engine – Powering America’s Exploration of Deep Space: The Engines Behind NASA’s Space Launch System”, from <https://www.nasa.gov/reference/space-launch-system-rs-25-core-stage-engine/> (accessed 1/18/2024).
2. Aerojet Rocketdyne, 2024, “RS-25 Engine | L3Harris® Fast. Forward,” from <https://www.l3harris.com/all-capabilities/rs-25-engine> (accessed 1/28/2024).
3. National Aeronautics and Space Administration, 2020, “Fracture Control Requirements for Spaceflight Hardware,” NASA-STD-5019A w/Change 3 revalidation with editorial /administrative changes, from <https://standards.nasa.gov/standard/NASA/NASA-STD-5019> (accessed 1/29/24).
4. National Aeronautics and Space Administration, 2023, “Nondestructive Evaluation Requirements for Fracture-Critical Metallic Components,” NASA-STD-5009C, from <https://standards.nasa.gov/standard/NASA/NASA-STD-5009> (accessed 1/29/24).
5. Department of Defense (DOD), 2009, “Nondestructive Evaluation System Reliability Assessment,” MIL-HDBK-1823A.
6. Agresti, A., 1990, *Categorical Data Analysis*, 1st ed. Wiley Series in Probability and Mathematical Statistics, Wiley-Interscience, New York.
7. Casella, G., and R.L. Berger, 1990, *Statistical Inference*, 1st ed., Duxbury Press, Belmont, CA.
8. Hahn, G.J., and S. Shapiro, 1994, *Statistical Models in Engineering,* Wiley Classics Library, Wiley-Interscience, New York.