
Visual Assessments of Complex Multicolored Patterns

Juan Lin

Introduction

- Visual Assessment Criteria
 - Individual Colors
 - Pass/Fail/Borderline
 - Gray Scale Ratings
 - Overall Pattern
 - Role of “Key” Pattern
- Methodology
 - Standard Places Adjacent to Batch
 - Standard Overlapping Batch
 - Assessments Based on Key/No Key
- Results
- Conclusion

Appearance Effects



Methodology

- Comparison of Standard and batch overlapped in the key region against results obtained from the Natick expert assessor (P/F).
- Assessing each color in the Standard against that in batch in the key region and determining the visual difference using the Gray Scales.
- Assessing Standard adjacent to batch in the key region against results obtained from the Natick assessor (P/F).
- Correlating visual pass/fail assessments from observers at NCSU for both methods above with those obtained from the expert assessor at Natick Soldier Division Center.

Methodology

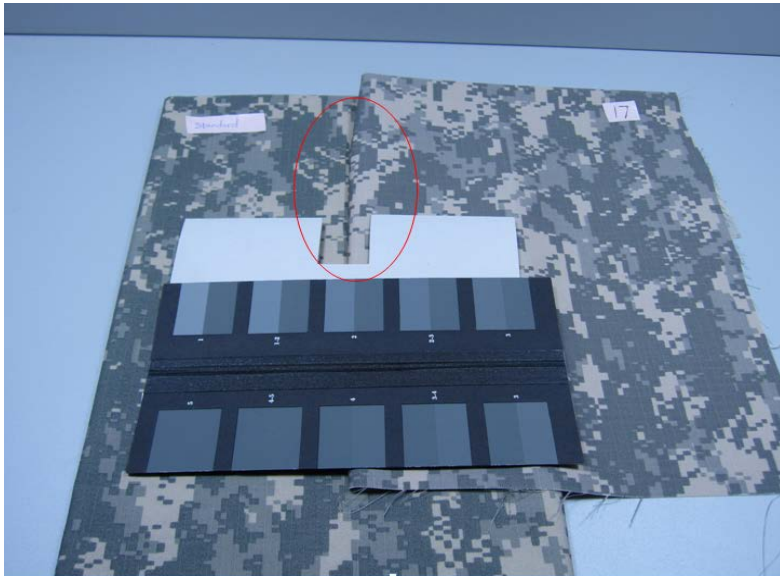
- Using the statistical function, STRESS, evaluating observers' consistency and repeatability in determining Pass/Fail assessments as well as in gray scale ratings.
- Determining the visual assessment protocol resulting in the smallest variability amongst observers and generating the closest agreement with results obtained from the expert assessor at Natick Soldier Division Center.

Methodology–Adjacent Visual Assessments

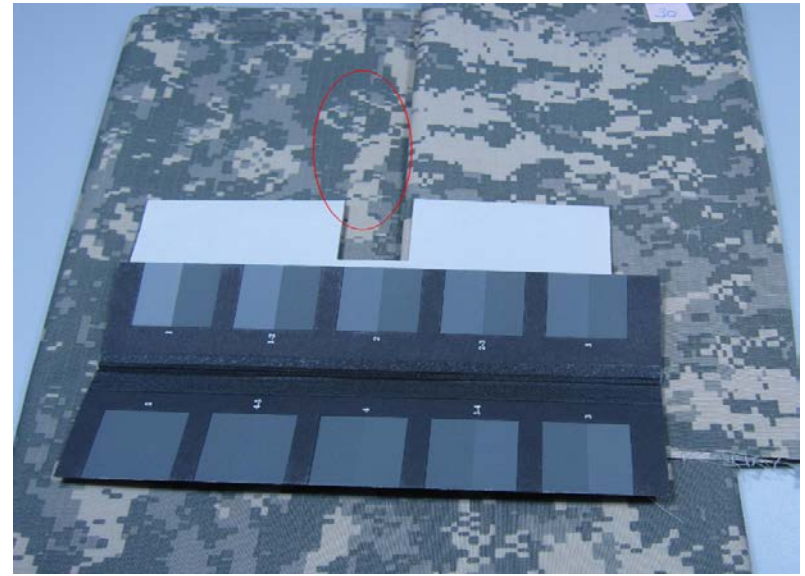


The KEY pattern identified with red ovals on the batch (right) and standard (left) samples placed adjacent to one another.

Overlapped Samples in Visual Assessments



Using the Key Pattern



With no Key Pattern

Visual Assessments of Complex Patterns

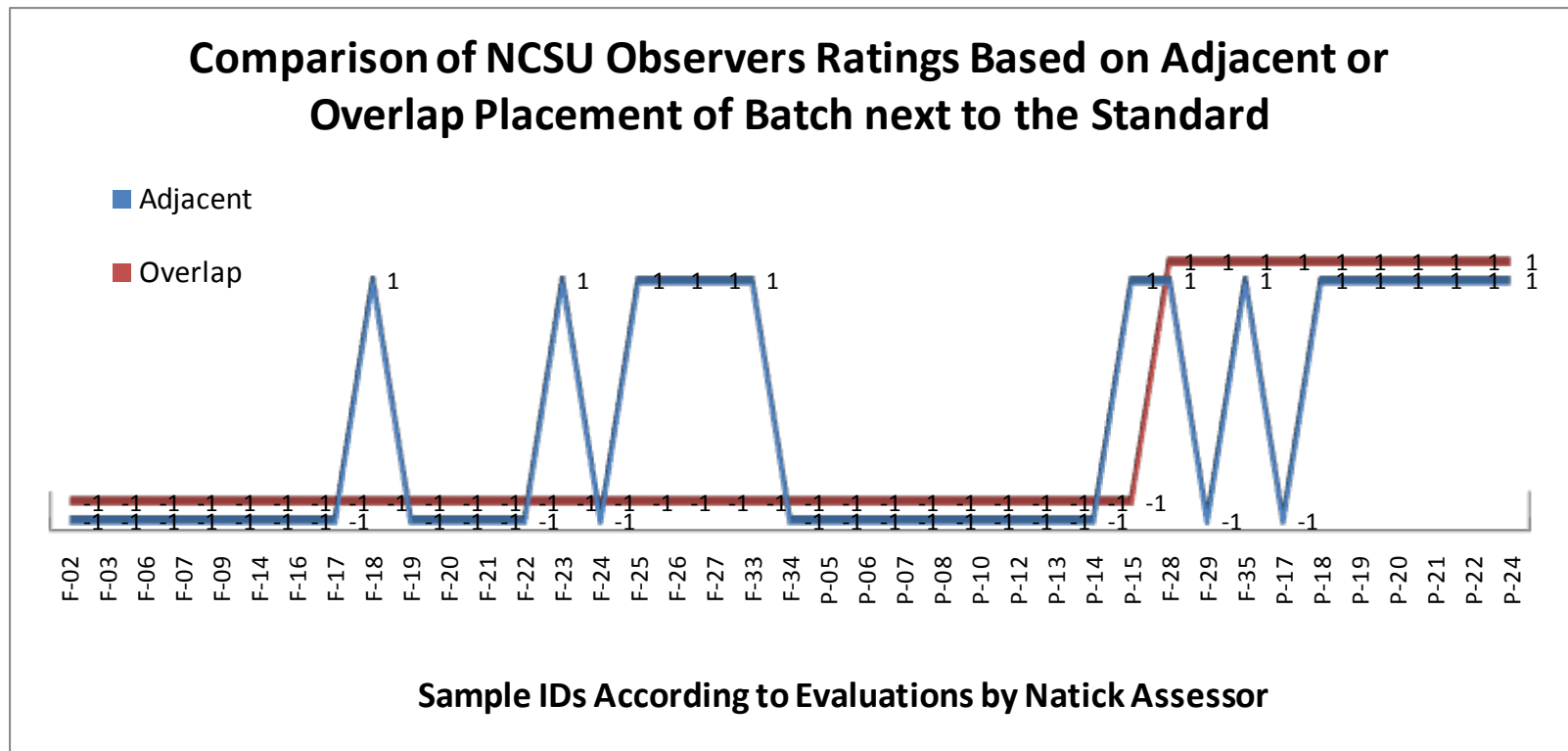


Geometry of Illumination/Viewing Used in Visual Assessments

Results and Discussions

- Role of Sample Placement
- Role of Assessment Criteria (P/F, GS)
- Role of Key Pattern
- Comparison of Mean Observer vs Natick

Ratings based on Adjacent or Overlap Placement of Batch next to the Standard (NCSU)



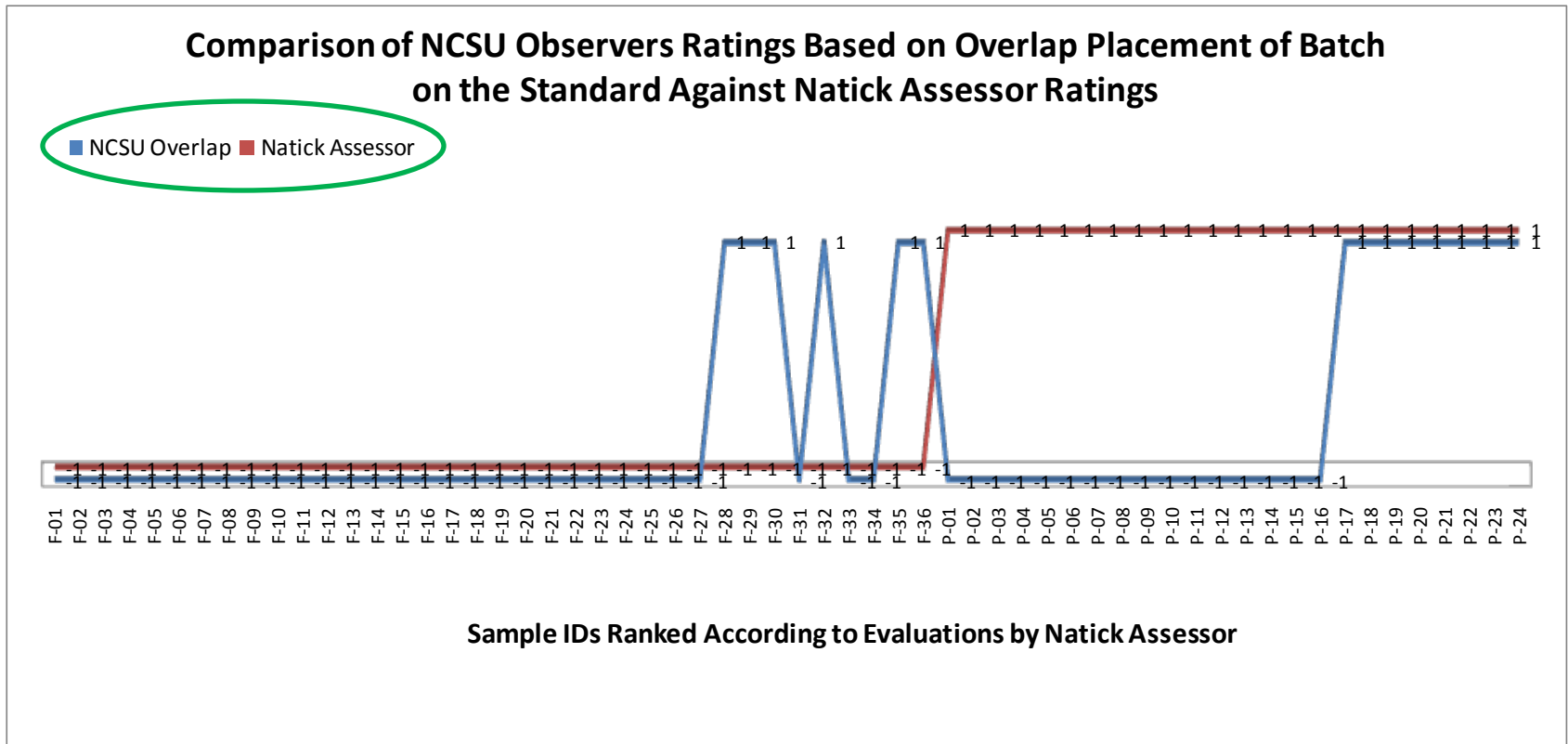
Comparison of results for assessments carried out by NCSU observers when batch samples with Key pattern were placed adjacent (blue solid line) to standard and those when they were laid on top of the standard (red solid line) during visual assessments. Value of 1 indicates Pass ratings and -1 indicates Fail ratings

Ratings based on Adjacent or Overlap Placement of Batch next to the Standard (NCSU)

Samples	Agreement Threshold (%)	Pass or Fail Samples Assessed Using Adjacent Principle	Pass or Fail Samples Assessed Using Overlapping Principle
Number PASS (24)	≥ 75	P-24 (1)	P-24 (1)
	50-75	P-15, P-18, P-19, P-20, P-21, P-22, F-18, F-23, F-25, F-26, F-27, F-28, F-33, F-35. (14)	P-17, P-18, P-19, P-20, P-21, P-22, P-23, F-27, F-28, F-29, F-30, F-31, F-32, F-33, F-34, F-35, F-36. (17)
Number FAIL (36)	≥ 75	P-05, P-06, P-12, P-14, F-02, F-03, F-06, F-07, F-09, F-12, F-21. (11)	P-01, P-02, P-03, P-05, F-01, F-02, F-03, F-05, F-06, F-07, F-08, F-09, F-10, F-11. (14)
	50-75	P-08, P-10, P-13, F-14, F-16, F-17, F-19, F-20, F-24, F-25, F-27, F-29, F-34. (13)	P-06, P-07, P-08, P-09, P-10, P-11, P-12, P-13, P-14, P-15, P-16, P-17, P-18, P-19, F-12, F-13, F-14, F-15, F-16, F-17, F-18, F-19, F-20, F-21, F-22, F-23, F-24, F-25, F-26, F-27, F-28, F-29, F-30, F-31, F-31, F-33, F-34. (37)

Comparison of Pass/Fail Samples Assessed According to Adjacent and Overlap Principles. Numbers in brackets indicate the total number of samples that meet the criteria in that group.

Ratings based on Overlap Placement of Batch on the Standard Against Natick Assessor Ratings



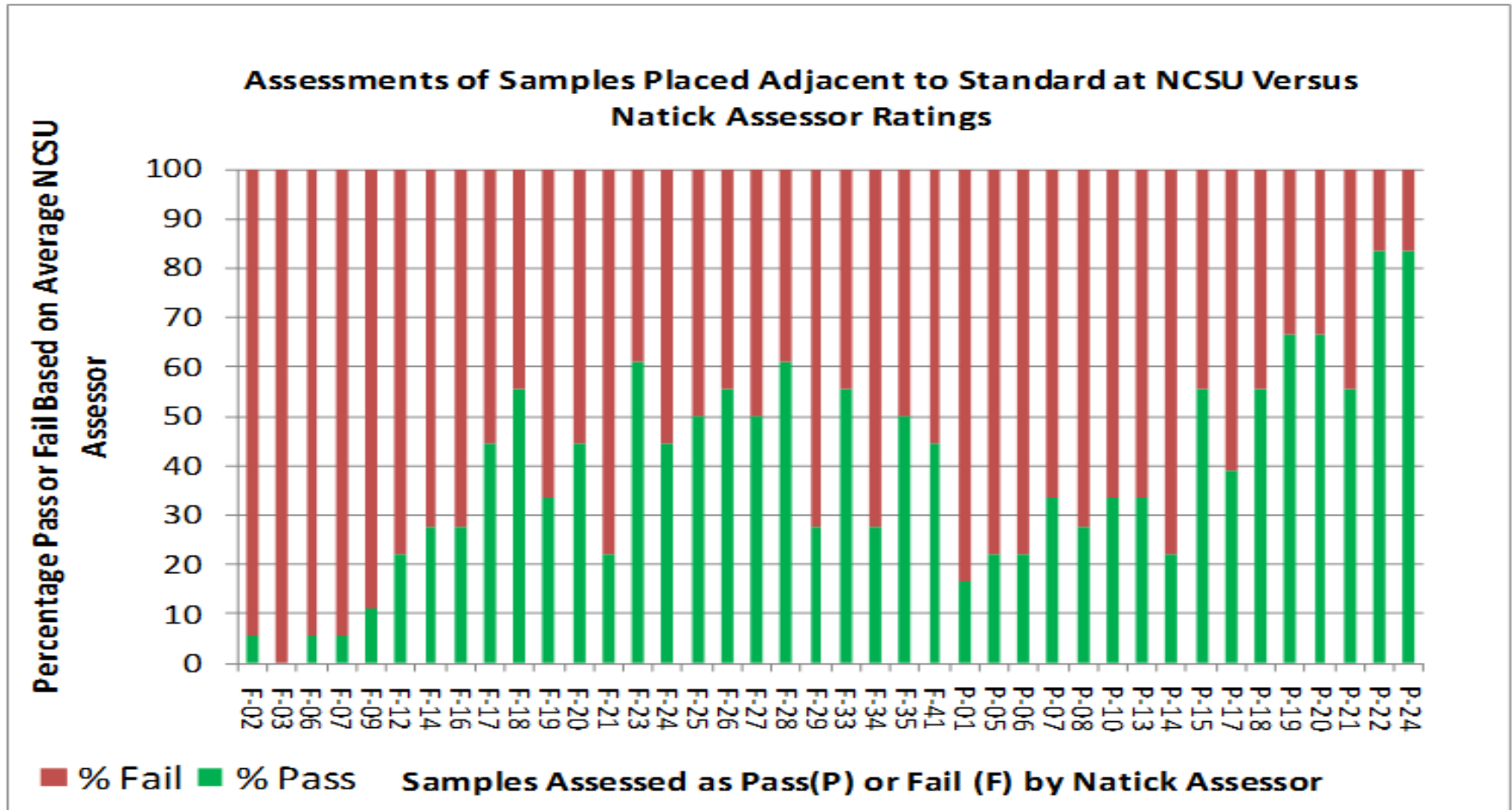
Comparison of results for assessments carried out by NCSU observers when batch samples were placed on top of the standard (blue solid line) and those evaluated in the same manner by Natick expert assessor (red solid line). Value of 1 indicates Pass ratings and -1 indicates Fail ratings

Comparison of P/F Results for Different Settings

%P/F Agreement	Natick	<u>Overlap</u> _NCSU	<u>Adjacent</u> _NCSU
Natick	100	63	57.5
<u>Overlap</u> _NCSU		100	77.5
<u>Adjacent</u> _NCSU			100

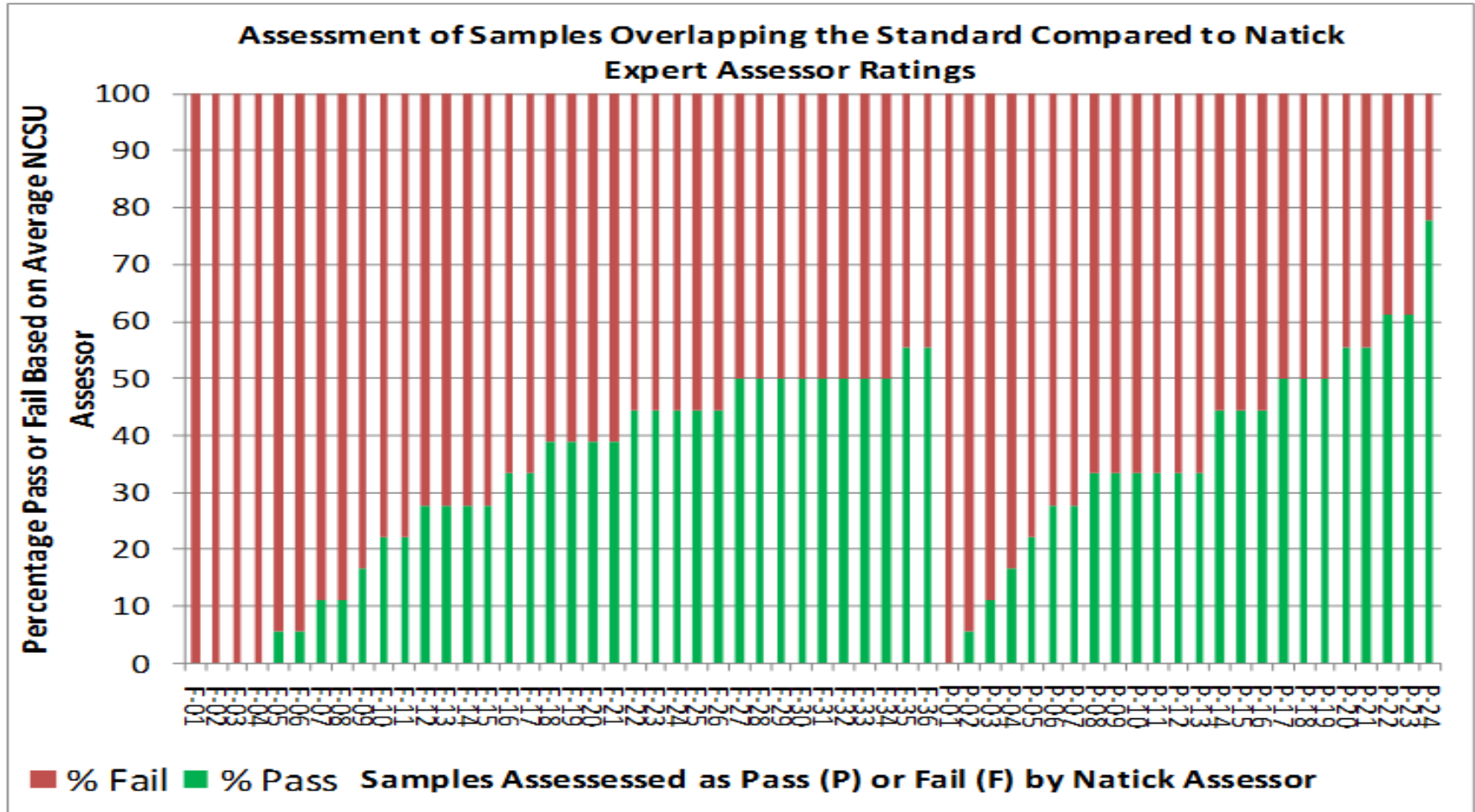
Agreement of overall assessment of samples in different occasions

Comparison of Samples Placed Adjacent to Standard at NCSU VS Natick



Overall % Pass/Fail of samples placed adjacent to the standard assessed by NCSU observers and compared to assessments done by expert assessor at Natick

Comparison of Samples Overlapping the Standard at NCSU VS Natick



Overall % Pass/Fail of samples placed on top of the standard and assessed by NCSU observers and compared to assessments done by expert assessor at Natick

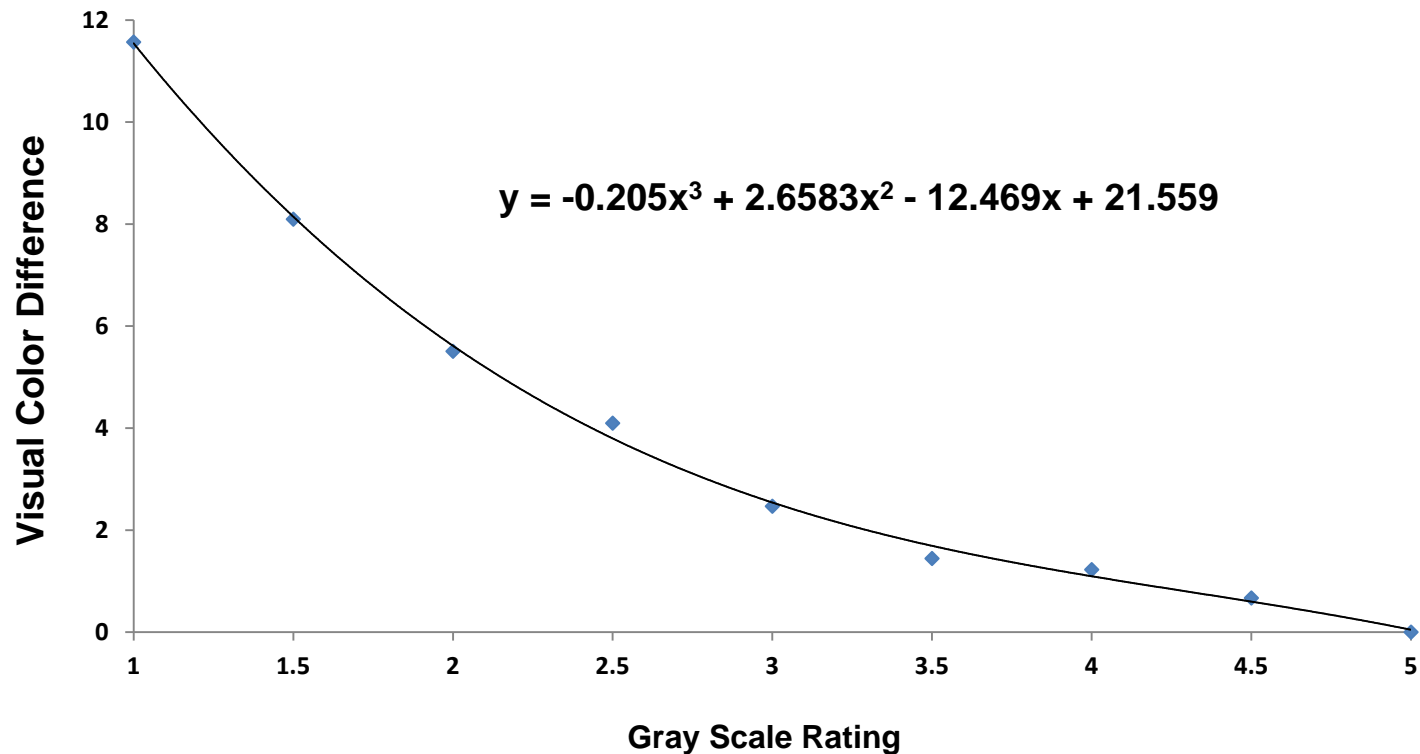
Gray Scale Transformations to Visual Difference

❖ Visual Assessment for Natick Project

Gray Scale Ratings \longrightarrow colorimetric data (D75)

Gray Scale	L^*	a	b	c	h	X	Y	Z	ΔE
5	41.41	-0.84	-1.12	1.40	233.28	11.35	12.12	14.57	0.00
4.5	41.94	-0.51	-1.35	1.44	249.29	11.72	12.47	15.07	0.67
4	42.61	-0.99	-0.95	1.37	223.92	12.09	12.90	14.23	1.22
3.5	42.82	-0.85	-1.43	1.67	239.53	12.21	13.04	15.08	1.44
3	43.87	-0.69	-1.13	1.32	238.54	12.90	13.76	16.49	2.47
2.5	45.49	-0.80	-1.40	1.61	240.52	13.97	14.91	17.98	4.10
2	46.91	-1.03	-1.30	1.66	231.65	14.91	15.95	19.19	5.51
1.5	49.50	-0.94	-1.30	1.61	234.04	16.86	18.01	21.61	8.10
1	52.97	-0.63	-1.11	1.28	240.68	19.76	21.03	25.08	11.57

Conversion of Gray Scale Ratings to VD



Conversion of Gray Scale Rating to Visual Differences

Observer Variability

Standardized Residual Sum of Squares Index (STRESS)

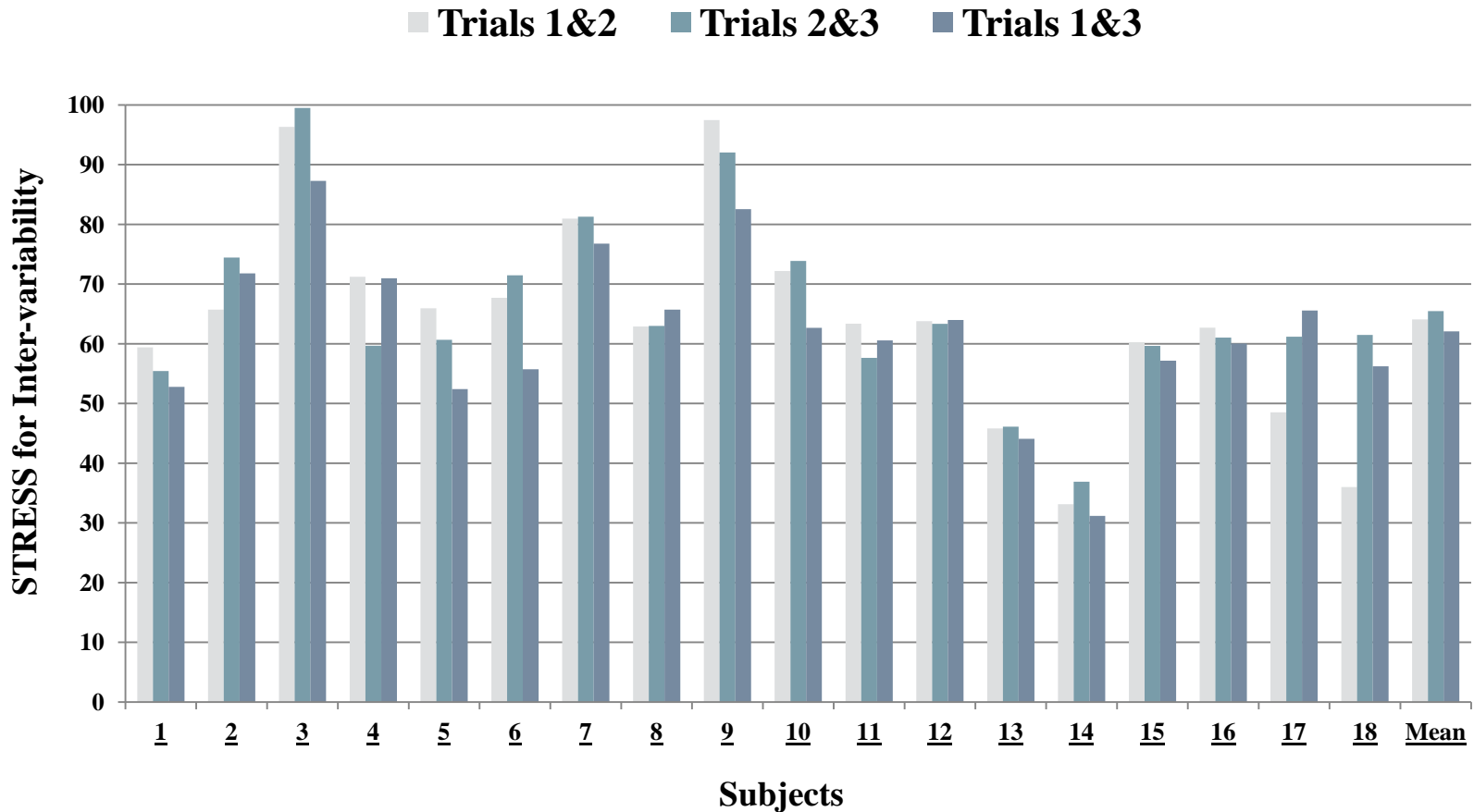
$$STRESS = 100 \sqrt{\frac{\sum (\Delta E_i - F_1 \Delta V_i)^2}{\sum F_1^2 \Delta V_i^2}}$$

$$F_1 = \frac{\sum \Delta E_i^2}{\sum \Delta E_i \Delta V_i}$$

ΔV_i is visually perceived color difference, ΔE_i is computed color difference.

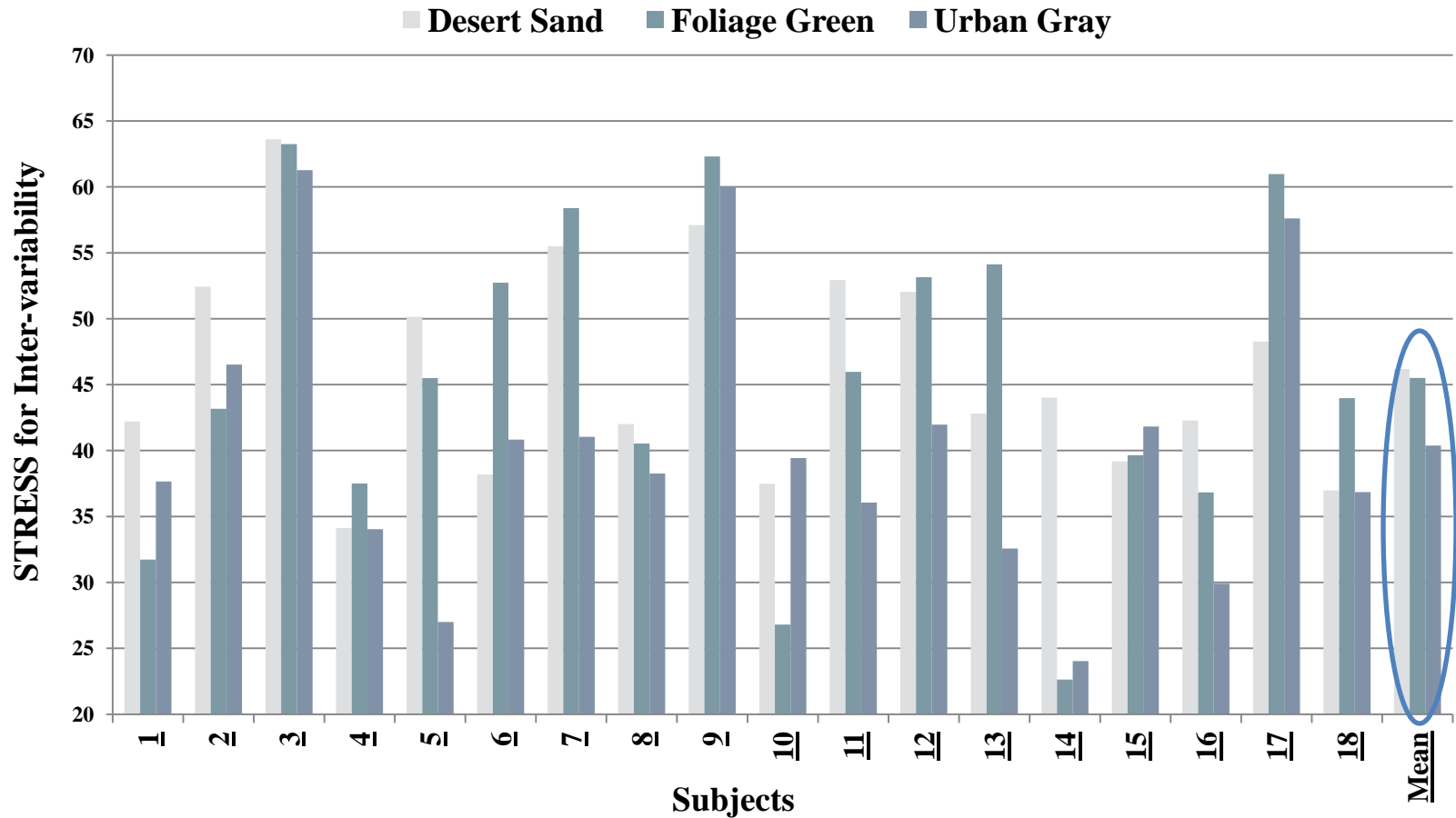
STRESS can be employed to determine observer variability in visual assessments.

Intra-observer variability using STRESS



STRESS for Intra-Observer Variability of Assessments of Urban Gray

Inter-observer Variability using STRESS



Conclusions

- Visual assessments conducted at NCSU have a modest agreement of 63% with those obtained from the expert assessor at Natick.
- The agreement is higher for samples that contain the key pattern and is markedly reduced when samples do not contain the key pattern.
- Recommend to use samples containing an identical section (preferably the repeat pattern such as the key) for use as a visual reference in evaluations.

Conclusions

- Visual assessments of colors by overlapping batch sample on the standard yield more reproducible results compared to assessments according to batch placed adjacent to the standard.
- Placement of a mask in the region of interest minimizes the effect of the surround on perception of color differences (reduced simultaneous contrast).

Conclusions

- Intra-observer variability STRESS in trials is relatively low.
- Inter-observer variability STRESS of NCSU observers consistent within the group and thus reasonably reproducible.
- STRESS can be used as a useful tool to screen observers prior to conducting full scale visual assessments.

Acknowledgments

- Observers
- Natick and US Army For Financial Support
- Thank you for your attention