

# Visual Diagnostics for Algorithmic Cartridge Case Comparisons

Joseph Zemmels, Heike Hofmann, Susan VanderPlas

## **Acknowledgements**

Thank you to everyone at the Roy J Carver High Resolution Microscopy Facility for collecting cartridge case scans.

### **Funding statement**

This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, University of Pennsylvania, Swarthmore College and University of Nebraska, Lincoln.

# Background

## Cartridge Case Comparisons

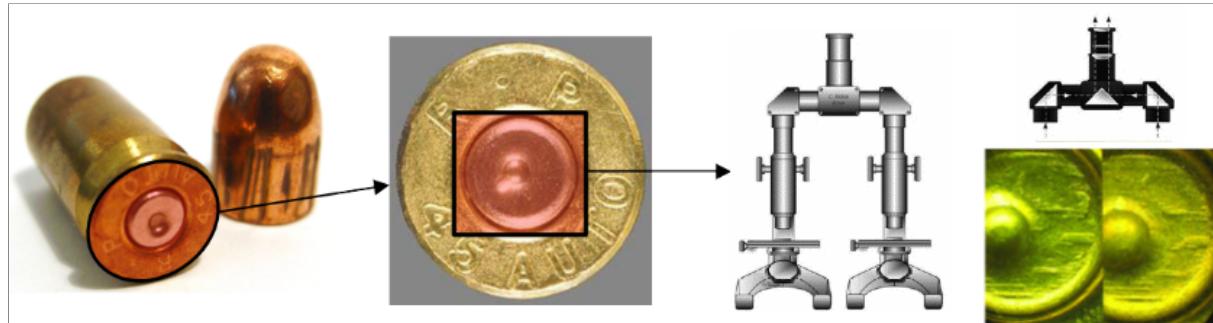
- Determine whether two cartridge cases were fired from the same firearm.



- **Cartridge Case:** metal casing containing primer, powder, and a projectile
- **Breech Face:** back wall of gun barrel
- **Breech Face Impressions:** markings left on cartridge case surface by the breech face during the firing process

## Current Practice

- Cartridge case recovered from crime scene vs. fired from suspect's firearm
- Place evidence under a comparison microscope for simultaneous viewing ([Thompson 2017](#))
- Assess the “agreement” of impressions on the two cartridge cases ([AFTE Criteria for Identification Committee 1992](#))



## Impression Comparison Algorithms

National Research Council ([2009](#)):

*“[T]he decision of a toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates”*

President's Council of Advisors on Science and Technology ([2016](#)):

*“A second - and more important - direction is (as with latent print analysis) to convert firearms analysis from a subjective method to an objective method. This would involve developing and testing image-analysis algorithms for comparing the similarity of tool marks on bullets [and cartridge cases].”*

We discuss the **Automatic Cartridge Evidence Scoring** (ACES) algorithm to compare 3D topographical images of cartridge cases

- Visual diagnostics aid in understanding what the algorithm does “under the hood.”

GIF source: <https://imgur.com/4CXf9BK>

# **Cartridge Case Comparison Algorithms**

## Ames I Study

- Baldwin et al. ([2014](#)) collected cartridge cases from 25 Ruger SR9 pistols
- Separated cartridge cases into quartets: 3 *known-match* + 1 *unknown source*
- *Match* if fired from the same firearm, *Non-match* if fired from different firearms
- 218 examiners tasked with determining whether the unknown cartridge case originated from the same pistol as the known-match cartridge cases
  - *True Positive* if a match is correctly classified, *True Negative* if non-match is correctly classified

	Match Conclusion	Non-match Conclusion	Inconclusive Conclusion	Total
<b>Ground-truth Match</b>	1,075	4	11	1,090
<b>Ground-truth Non-match</b>	22	1,421	735 + 2*	2,180
<b>True Positive (%)</b>	99.6	<b>True Negative (%)</b>	65.2	<b>Overall Inconclusives (%)</b>
				22.9

\*Two non-match comparisons were deemed “unsuitable for comparison”

## **Cartridge Case Data**

- 3D topographic images using Cadre™ TopMatch scanner from Roy J Carver High Resolution Microscopy Facility
- **x3p** file contains surface measurements at lateral resolution of 1.8 micrometers (“microns”) per pixel

## Cartridge Case Comparison Algorithms

Obtain an objective measure of similarity between two cartridge cases

- **Step 1:** Independently *pre-process* scans to isolate breech face impressions
- **Step 2:** Compare two cartridge cases to extract a set of numerical features that distinguish between matches vs. non-matches
- **Step 3:** Combine numerical features into a single similarity score (e.g., similarity score between 0 and 1)

Examiner takes similarity score into account during an examination

Challenging to know how/when these steps work correctly

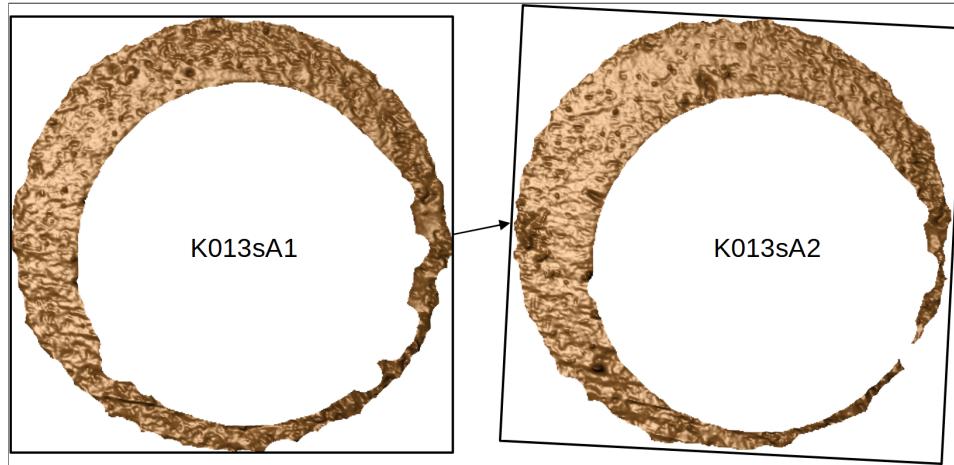
## **Step 1: Pre-process**

Isolate region in scan that consistently contains breech face impressions

***How do we know when a scan is adequately pre-processed?***

## Step 2: Compare Full Scans

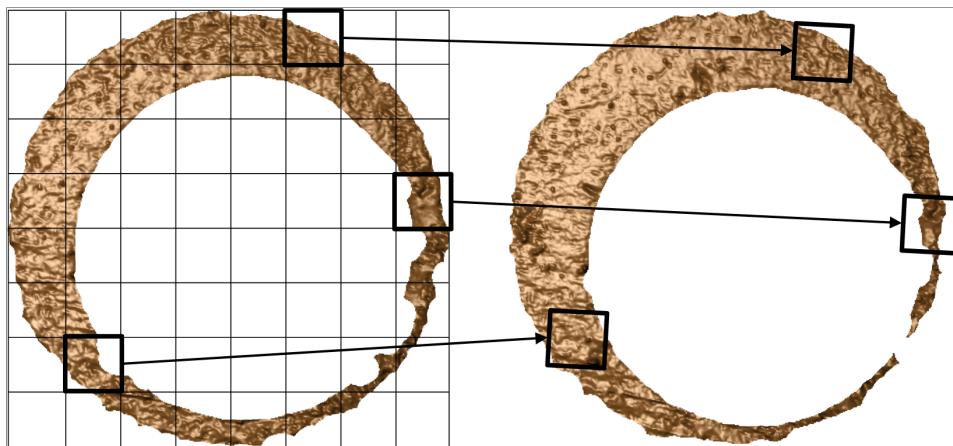
- *Registration:* Determine rotation and translation to align two scans



- *Cross-correlation function (CCF)* measures similarity between scans
  - Choose the rotation/translation that maximizes the CCF

## Step 2: Compare Cells

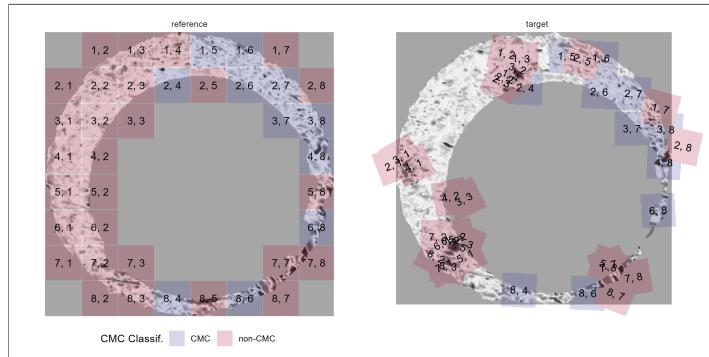
- Split one scan into a grid of cells that are each registered to the other scan ([Song 2013](#))
- For a matching pair, we assume that cells will agree on the same rotation & translation



*Why does the algorithm “choose” a particular registration?*

## Step 3: Score

- Measure of similarity for two cartridge cases
  - Maximized CCF (0.27 in example below) ([Vorburger et al. 2007; Tai and Eddy 2018](#))
  - Congruent Matching Cells (11 CMCs in example below) ([Song 2013](#))



- Our approach: similarity score between 0 and 1 using a statistical model

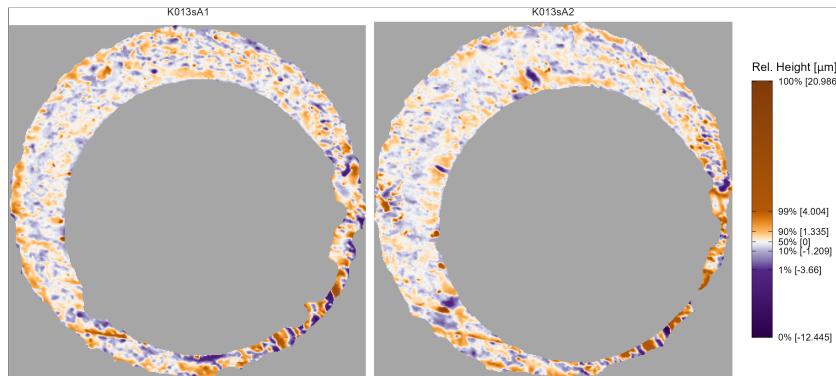
***What factors influence the final similarity score?***

# **Visual Diagnostics**

## Visual Diagnostics for Algorithms

- A number of questions arise out of using comparison algorithms
  - *How do we know when a scan is adequately pre-processed?*
  - *Why does the algorithm “choose” a particular registration?*
  - *What factors influence the final similarity score?*
- We wanted to create tools to address these questions
  - Well-constructed visuals are intuitive and persuasive
  - Useful for both researchers and practitioners to understand the algorithm’s behavior

## X3P Plot

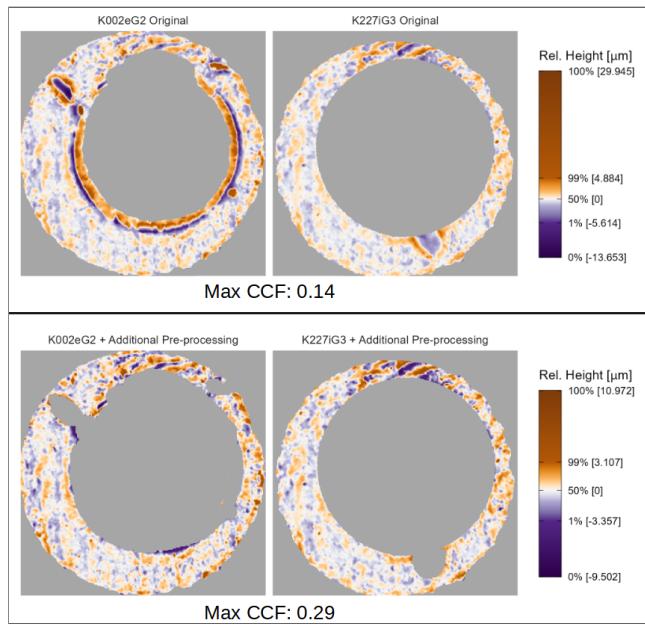


- Emphasizes extreme values in scan that may need to be removed during pre-processing
- Allows for comparison of multiple scans on the same color scheme
- Map quantiles of surface values to a divergent color scheme

Min (0°) #2D004B	1° #542788	2.5° #8073AC	10° #B2ABD2	25° #D8DAEB	50° #F7F7F7	75° #FEE0B6	90° #FDB863	97.5° #E08214	99° #B35806	Max (100°) #7F3B08
---------------------	---------------	-----------------	----------------	----------------	----------------	----------------	----------------	------------------	----------------	-----------------------

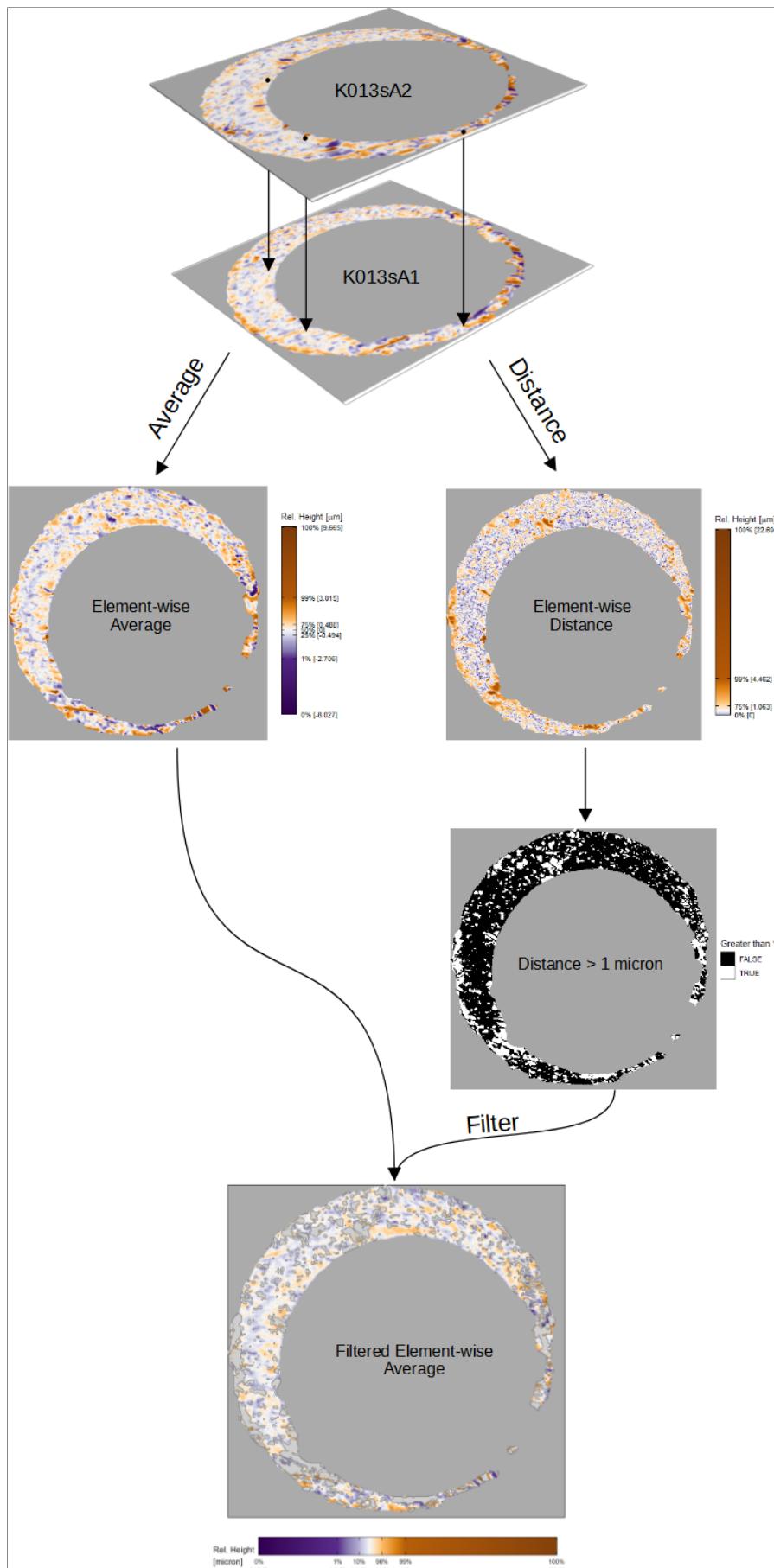
## X3P Plot Pre-processing Example

- Useful for diagnosing when scans need additional pre-processing

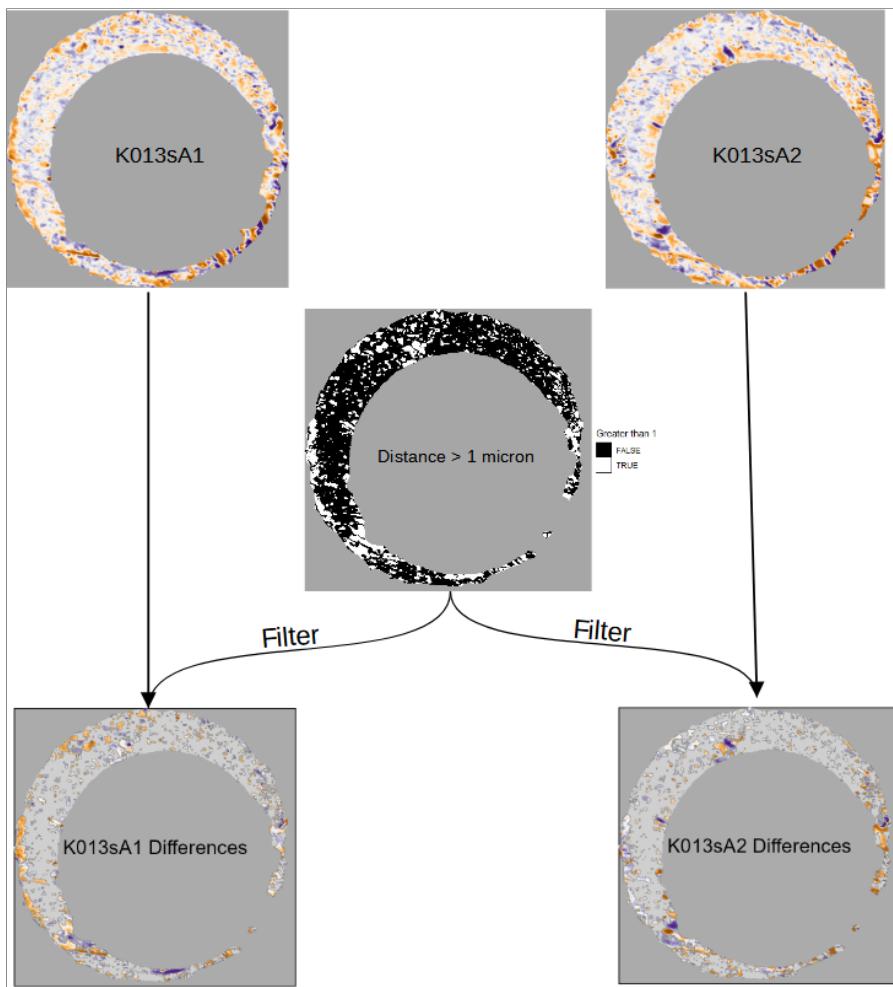


## Comparison Plot

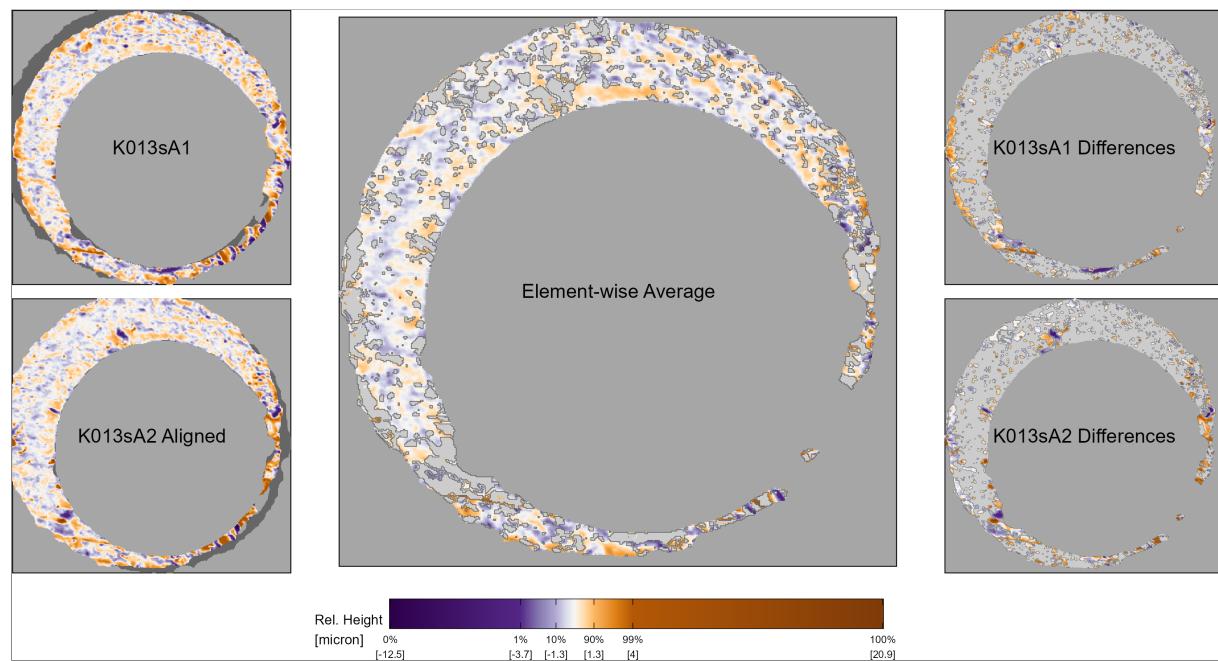
- Separate aligned scans into similarities and differences
- Useful for understanding a registration
- *Similarities:* Element-wise average between two scans after filtering elements that are less than 1 micron apart



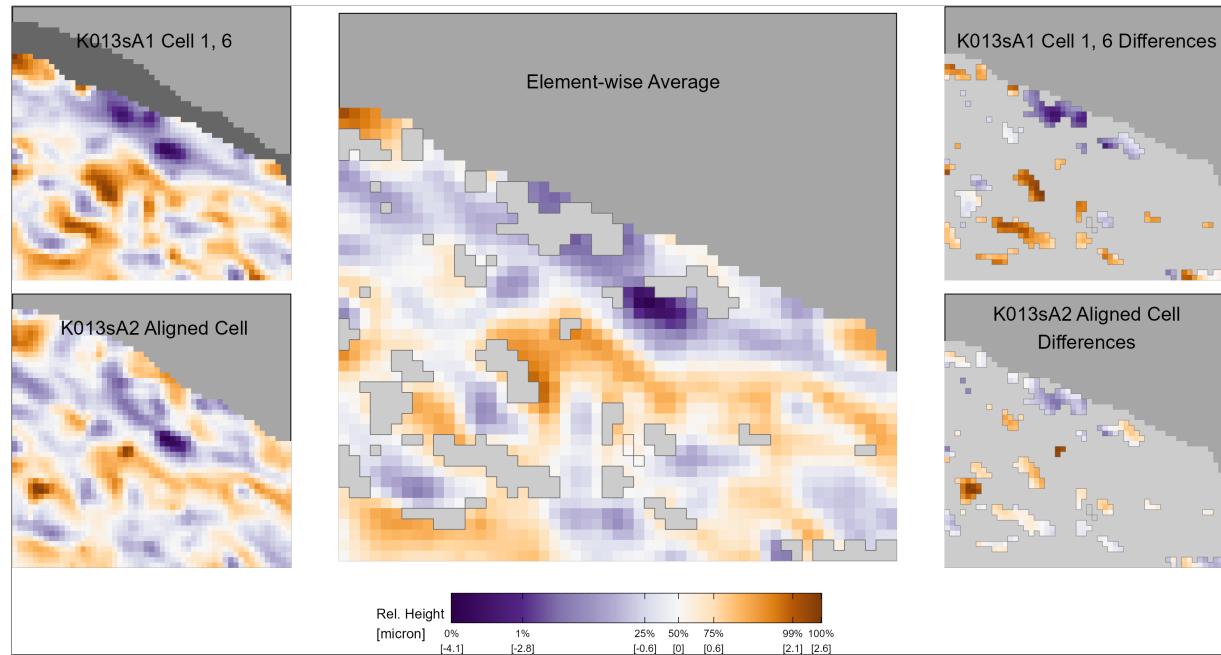
- *Differences:* Elements of both scans that are at least 1 micron apart



## Full Scan Comparison Plot



# Cell Comparison Plot



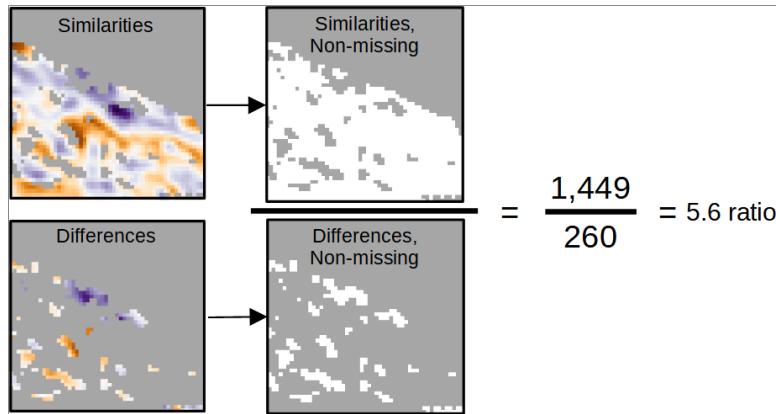
## Translating Visuals to Statistics

- Translate qualitative observations made about the visual diagnostics into complementary numerical statistics
- Useful to quantify what our intuition says should be true for (non-)matching scans
- For a matching cartridge case pair...
  1. There should be (many) more similarities than differences
  2. The different regions should be relatively small
  3. The surface values of the different regions should follow similar trends
- Statistics are useful for justifying/predicting the behavior of the algorithm

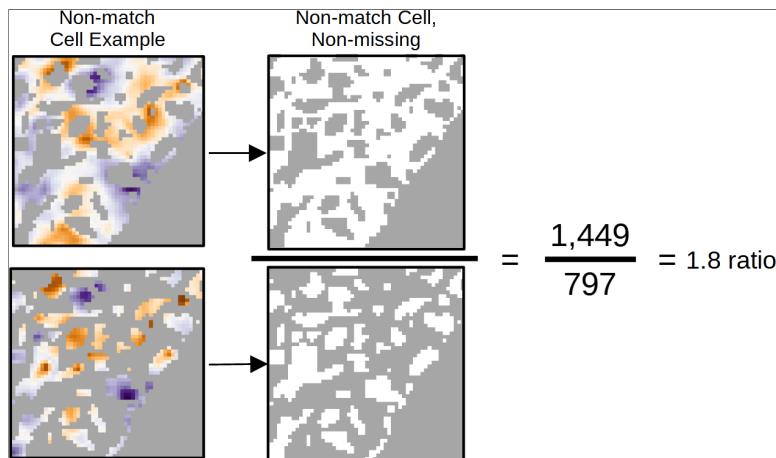
## Similarities vs. Differences Ratio

1. There should be more similarities than differences

*Ratio between number of similar vs. different observations*



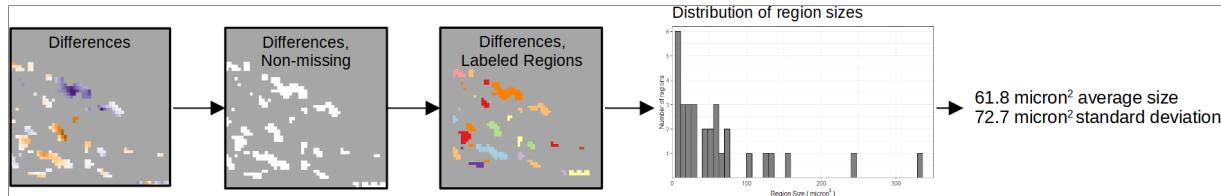
Compare to a non-match cell comparison:



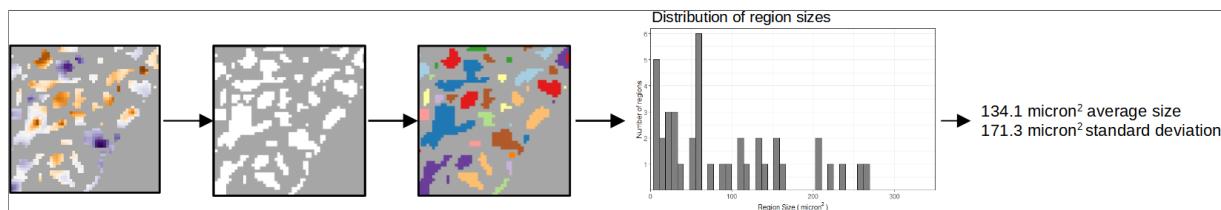
## Different Region Size

2. The different regions should be relatively small

### *Size of the different regions*



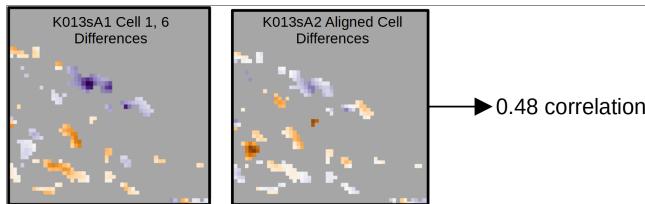
Compare to a non-match cell comparison:



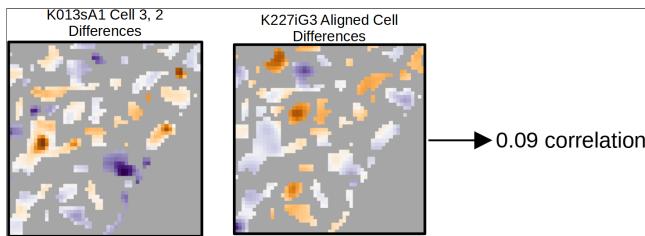
## Different Region Correlation

3. The surface values of the different regions should follow similar trends

*Correlation between the different regions of the two scans*

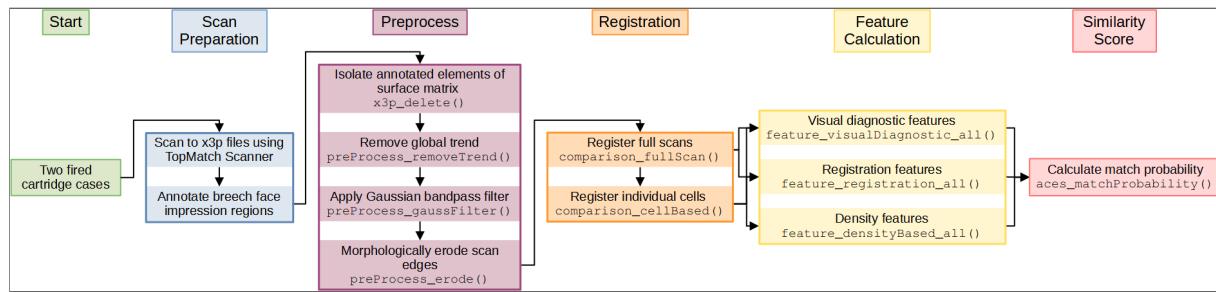


Compare to a non-match cell comparison:



# **Automatic Cartridge Evidence Scoring (ACES) Algorithm**

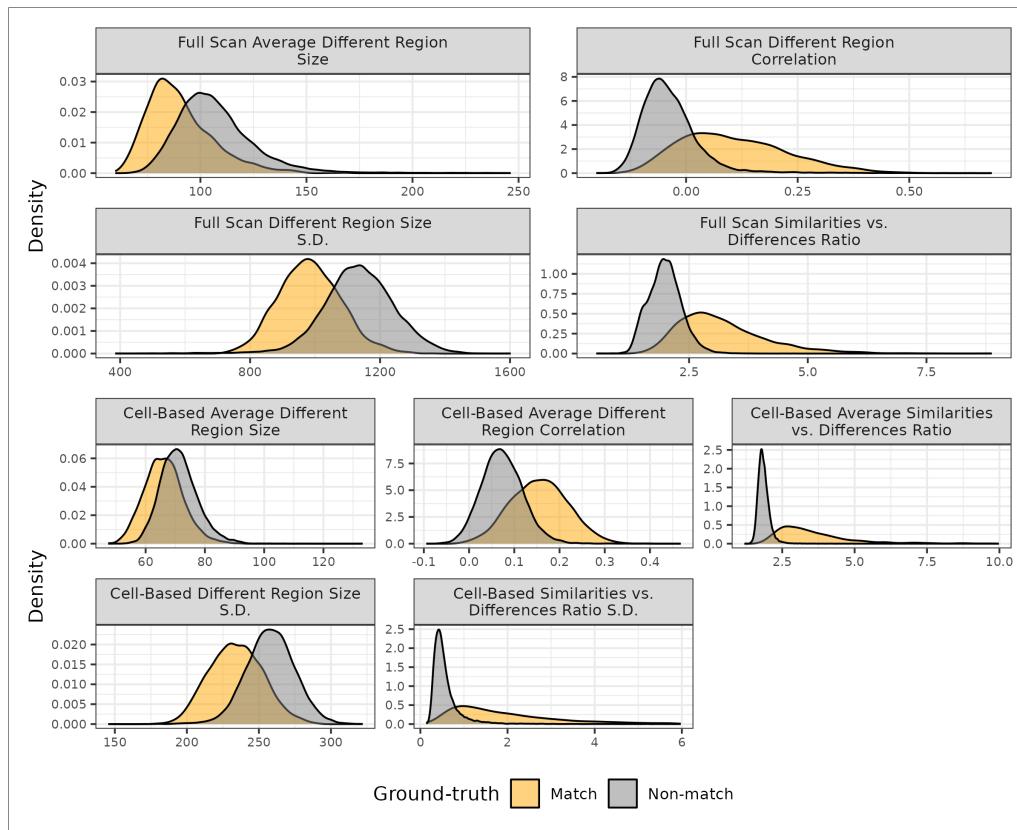
# Automatic Cartridge Evidence Scoring



- Comparison algorithm that pre-processes, compares, and scores two cartridge case scans
- Computes 19 numerical features for each cartridge case pair
- Computes similarity score between 0 and 1 for a cartridge case pair using trained statistical model

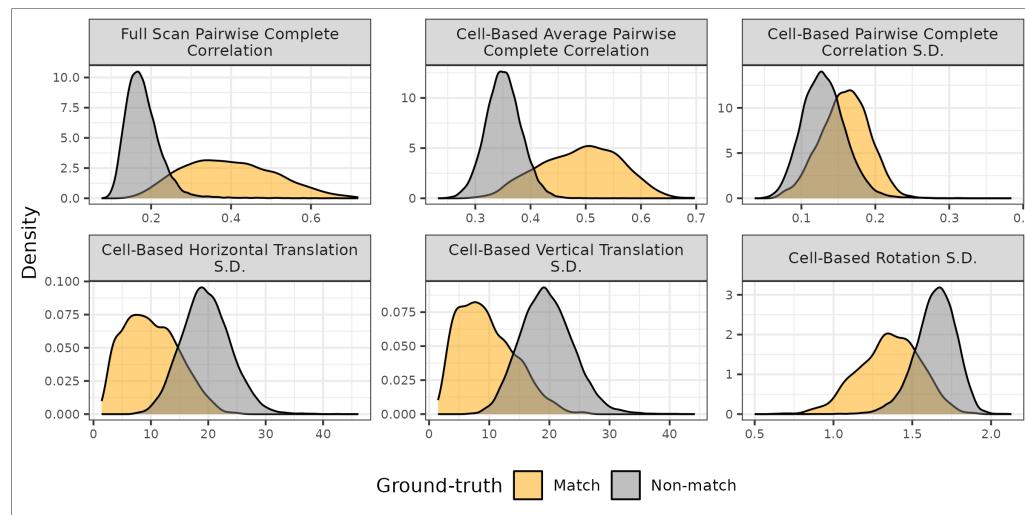
# Visual Diagnostic Features

- Use visual diagnostic statistics discussed earlier as numerical features
- Features:
  - From the full scan comparison:
    - Similarities vs. differences ratio
    - Average and standard deviation of different region sizes
    - Different region correlation
  - From cell-based comparison:
    - Average and standard deviation of similarities vs. differences ratios
    - Average and standard deviation of different region sizes
    - Average different region correlation



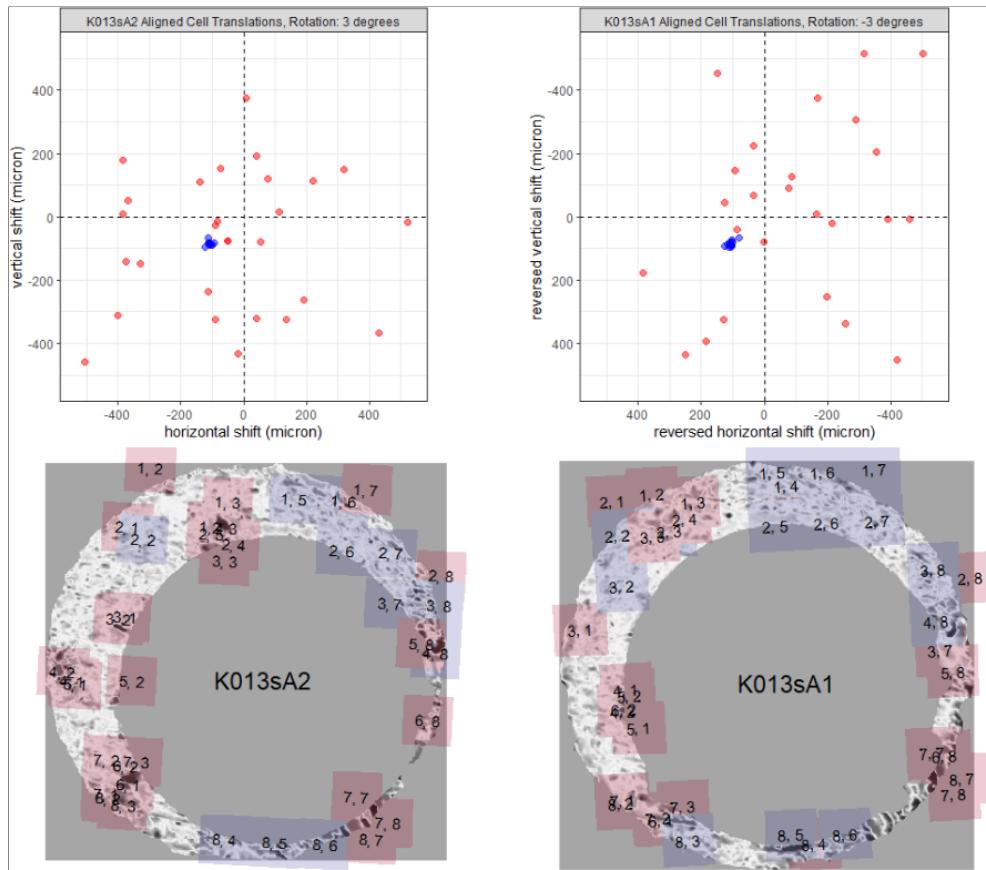
## Registration-based Features

- For a matching cartridge case pair...
  - Correlation should be large at the full scan *and* cell levels
  - Cells should “agree” on a particular registration
- Compute summary statistics of full-scan and cell-based registration results
- Features:
  - Correlation from full scan comparison
  - Mean and standard deviation of correlations from cell comparisons
  - Standard deviation of cell-based registration values (horizontal/vertical translations & rotation)

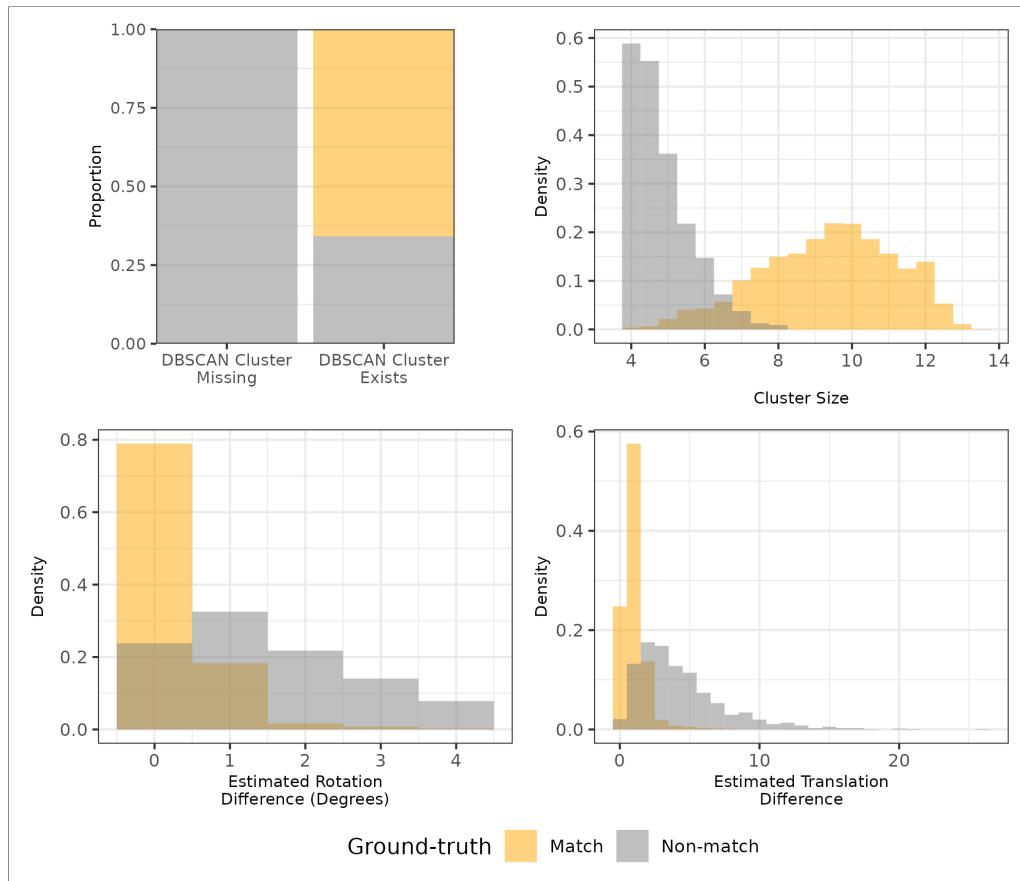


## Density-based Features

- For a matching cartridge case pair...
  - Cells should “agree” on a particular registration
  - The estimated registrations between the two comparison directions should be opposites
- Apply Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to the cell-based registration results ([Ester et al. 1996](#); [Zhang et al. 2021](#))



- Features:
  - DBSCAN cluster indicator
  - Average DBSCAN cluster size
  - Absolute sum of density-estimated rotations
  - Root sum of squares of the cluster-estimated translations



## ACES Statistical Model

- Compute 19 features for each pairwise comparison
- Use 510 cartridge cases from Baldwin et al. (2014) to fit a *logistic regression* classifier
- Train random logistic regression using 21,945 pairwise comparisons from 210 scans
  - Classify pairs as a “match” or “non-match” based on similarity score
  - Explore two optimization criteria:
    - Model that minimizes overall error
    - Model that balances true positive and true negative rates
- Test model on 44,850 pairwise comparisons from 300 scans
  - Compute true positive and true negative rates for each model
  - Consider distributions of similarity scores for truly matching and non-matching pairs

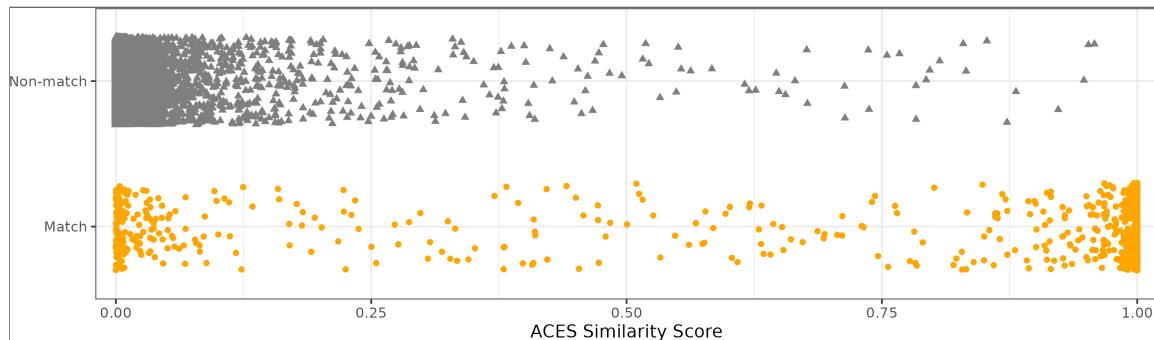
## Test Classification Results

Source	True Pos. (%)	True Neg. (%)	Overall Inconcl. (%)	Overall Acc. (%)
ACES, Min. Error	92.3	99.9	0.0	99.4
ACES, Balanced TP/TN	95.7	98.1	0.0	97.9
Ames I	99.6	65.2	22.9	

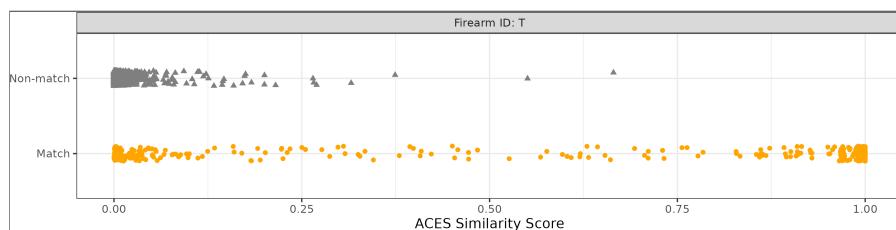
- Ames I ([Baldwin et al. 2014](#)) compared quartets (3 to 1) and considered inconclusives
- Class imbalance in test data: 3,081 match vs. 41,769 non-match comparisons
- The “Balanced TP/TN” model was selected based on the training data. The test data classifications aren’t guaranteed to also be balanced.

## Similarity Score Distributions

- We consider classification accuracy as a means of selecting/comparing models.
- In practice, the examiner would use the similarity score as part of their examination.



- Matching comparisons from Firearm T cartridge cases tend to have lower similarity scores:



# **Conclusions**

## Conclusions & Future Work

- Automatic comparison algorithms are useful for obtaining numerical measures of similarity for two pieces of evidence
- Visual diagnostics help explain the inner mechanisms of comparison algorithms
- Our visual diagnostic tools aid in understanding each step of a cartridge case comparison algorithm
  - Also useful by themselves to visually compare cartridge case evidence
- The Automatic Cartridge Evidence Scoring (ACES) algorithm shows promise at measuring the similarity between cartridge cases
- Develop free, open source software to implement visual diagnostics & ACES
  - We train our model on 10 firearms, all with the same make/model and ammunition
  - Need additional “stress tests” (different ammunition/firearms, degradation, etc.)

## Thank You!

Thank you all for joining me. On the slide are additional links to resources in case you are interested in learning more about the tools I discussed today

- **impressions** R package for visual diagnostics
  - <https://jzemmels.github.io/impressions/>
- **scored** R package for ACES algorithm
  - <https://jzemmels.github.io/scored/>
- **cartridgeInvestigatR** interactive web application
  - <https://csafe.shinyapps.io/cartridgeInvestigatR/>

## References

- AFTE Criteria for Identification Committee. 1992. "Theory of Identification, Range Striae Comparison Reports and Modified Glossary Definitions." *AFTE Journal* 24 (3): 336–40.
- Baldwin, David P, Stanley J Bajic, Max Morris, and Daniel Zamzow. 2014. "A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons." Fort Belvoir, VA: Ames Lab IA, Performing; Defense Technical Information Center. <https://doi.org/10.21236/ADA611807>.
- Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–31. KDD'96. Portland, Oregon: AAAI Press.
- National Research Council. 2009. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, D.C.: The National Academies Press.
- President's Council of Advisors on Science and Technology. 2016. "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods." Executive Office of The President's Council of Advisors on Science; Technology, Washington DC.
- Song, John. 2013. "Proposed 'NIST Ballistics Identification System (NBIS)' Based on 3d Topography Measurements on Correlation Cells." *American Firearm and Tool Mark Examiners Journal* 45 (2): 11. [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=910868](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=910868).
- Tai, Xiao Hui, and William F. Eddy. 2018. "A Fully Automatic Method for Comparing Cartridge Case Images," *Journal of Forensic Sciences* 63 (2): 440–48. <http://doi.wiley.com/10.1111/1556-4029.13577>.
- Thompson, Robert. 2017. *Firearm Identification in the Forensic Science Laboratory*. National District Attorneys Association. <https://doi.org/10.13140/RG.2.2.16250.59846>.
- Vorburger, T V, J H Yen, B Bachrach, T B Renegar, J J Filliben, L Ma, H G Rhee, et al. 2007. "Surface Topography Analysis for a Feasibility Assessment of a National Ballistics Imaging Database." NIST IR 7362. Gaithersburg, MD: National Institute of Standards; Technology. <https://doi.org/10.6028/NIST.IR.7362>.

Zhang, Hao, Jialing Zhu, Rongjing Hong, Hua Wang, Fuzhong Sun, and Anup Malik. 2021. "Convergence-Improved Congruent Matching Cells (CMC) Method for Firing Pin Impression Comparison." *Journal of Forensic Sciences* 66 (2): 571–82. <https://doi.org/10.1111/1556-4029.14634>.

## Appendix: Firearm-wise Probabilities

- Specific firearms in the test set tend to have lower associated similarity score for matching comparisons

