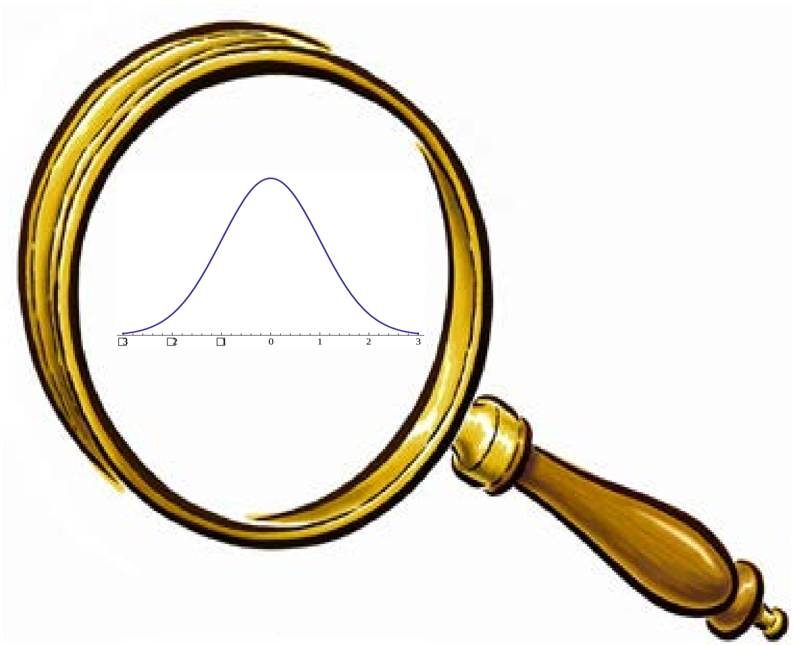
Approaches to Physical Evidence Research In the Forensic Sciences



Petraco Group
City University of New York, John Jay College

Outline

- Introduction
- Lessons learned from research and training in forensic science
- Recent alumni careers
- A bit about our research products:
    - Statistical tools implemented
- Directions for the future from lessons learned

Raising Standards with Data and Statistics

- DNA profiling the most successful application of statistics in forensic science.
  - Responsible for current interest in “raising standards” of other branches in forensics.
- No protocols for the application of statistics to physical evidence.
  - Our goal: **application of objective, numerical computational pattern comparison to physical evidence**

Learned Elements of Successful Research in Forensic Science

- My group has been involved with research and training in quantitative forensic science for about a decade.
- There must be a ***synergy between*** academics and practitioners!
    - Practitioners have specialized knowledge of:
        1. Technical problems in their fields
            - ***Don't dismiss their opinions*** for improving practice!
            - Successful research products will have been borne out by treating practitioners as equal partners
                - Co-PIs
                - "Test subjects" on proposed research products

Learned Elements of Successful Research in Forensic Science

- Practitioners have specialized knowledge of:

2. Intricacies of working within the criminal justice system
    - Academics are largely self-managed. ***Practitioners must deal with attorneys/judges/court systems***.

3. Intricacies of implementing new technologies/S.O.Ps
    - ***Lab's have tight budgets***: Consider equipment costs for purchase/maintenance/training
    - ***Realistic expectation*** for their staff's: knowledge/training/capacities
        - Will users need a PhD to do this??
        - User interfaces are important!!

Learned Elements of Successful Research in Forensic Science

- Successful University research programs will ***ultimately involve***:

1. ***"Well oiled" research teams of academics and practitioners*** with a ***proven dedication*** to research in forensic science.

2. Track-records of producing ***research products related to what was proposed*** to win the grant.

Learned Elements of Successful Research in Forensic Science

- Funding sources should ***consider***:

1. Significant ***PhD/D-Fos program start-up money*** for
    - Curriculum development
    - Teaching time/classroom space
    - Research space
    - Travel for dissemination of research products/training
    - Equipment purchase/maintenance costs
        - ***Should be close monitoring*** of nascent PhD/D-Fos program ***by funding source***.
        - Top and middle ***administration*** must be involved and ***supportive for the long-term*** (10-15 years).

Learned Elements of Successful Research in Forensic Science

- Funding sources should ***<u>consider</u>***:

2. ***<u>Accessibility/outreach</u>*** of PhD/D-Fos program
   - Research in forensic science must me disseminated!!
     - Is the ***<u>nearest airport</u>*** to campus small and two hours away?
     - Does campus have access to short term ***<u>living facilities</u>***?
       - Specialists to come and teach.
       - Practitioner to come and take non-matric classes.
     - What are the ***<u>dedicated facilities for training</u>*** on campus?

   - Proximity to federal/state/local forensic laboratories and large metropolitan areas
     - Sources of ***<u>collaboration</u>*** and ***<u>external opportunities</u>***

Some of the Recent Career Tracks of John Jay Alumni

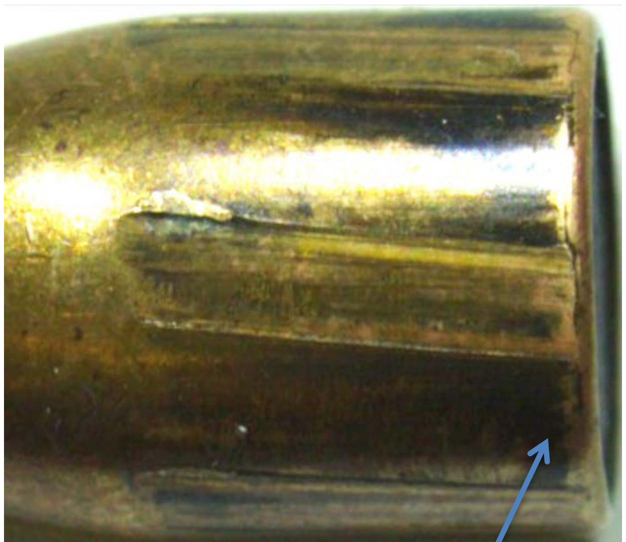
- **Federal, state and local public Crime laboratories**
  - Armed Forces DNA Identification Laboratory, Dover, DE; Center of Forensic Sciences, Toronto, Canada
  - Contra Costa County Sheriff’s Department Forensic Division, CA
  - Federal Bureau of Investigation Laboratory Division, Quantico, VA
  - Los Angeles Police Department, Scientific Investigation Division, LA, CA
  - Los Angeles Sheriff’s Department Laboratory, CA
  - Nassau County Crime Lab, East Meadow, NY
  - New Jersey State Police, Office of Forensic Sciences, West Trenton, NJ
  - New York City Office of Chief Medical Examiner, NY; New York City Police Department Laboratory, Jamaica, NY
  - Office of the Medical Examiner San Francisco, CA; Orange County Sheriff Department, Santa Ana, CA
  - San Diego Police Department Laboratory, San Diego, CA
  - Suffolk County Crime Lab, Hauppauge, NY
  - United States Drug Enforcement Agency, Laboratory Division NY, NY
  - United States Drug Enforcement Agency, Laboratory Division SF, CA; U.S. Postal Inspection Service, Forensic Laboratory, Dulles, VA
  - Washington DC Department of Forensic Sciences, DC
- **Private Companies**
  - Antech Diagnostics, New Hyde Park, NY; Bode Technology, Lorton, VA
  - Purdue Pharma, Totowa, NJ; Microtrace, Elgin, IL
  - Quest Diagnostics, Madison, NJ
  - Smiths Detection, Edgewood, MD; NMS Labs, Willow Grove, PA
  - Core Pharma, Middlesex, NJ
- **Academic and Research Institutions**
  - Central Police University of Taiwan
  - California State University, Los Angeles, CA
  - Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
  - John Jay College, New York, NY
  - Hawaii Chaminade University, Honolulu, HI; McMaster University, Hamilton, ON
  - St. Xaviers College, Bombay, India
  - Penn State, College Station, PA
  - University of New Haven, New Haven, CT
  - University of West Virginia, Morgantown, WV
  - University of Toronto, Canada
  - Weill Cornell Medical College, New York, NY

Our Research products:
Quantitative Criminalistics

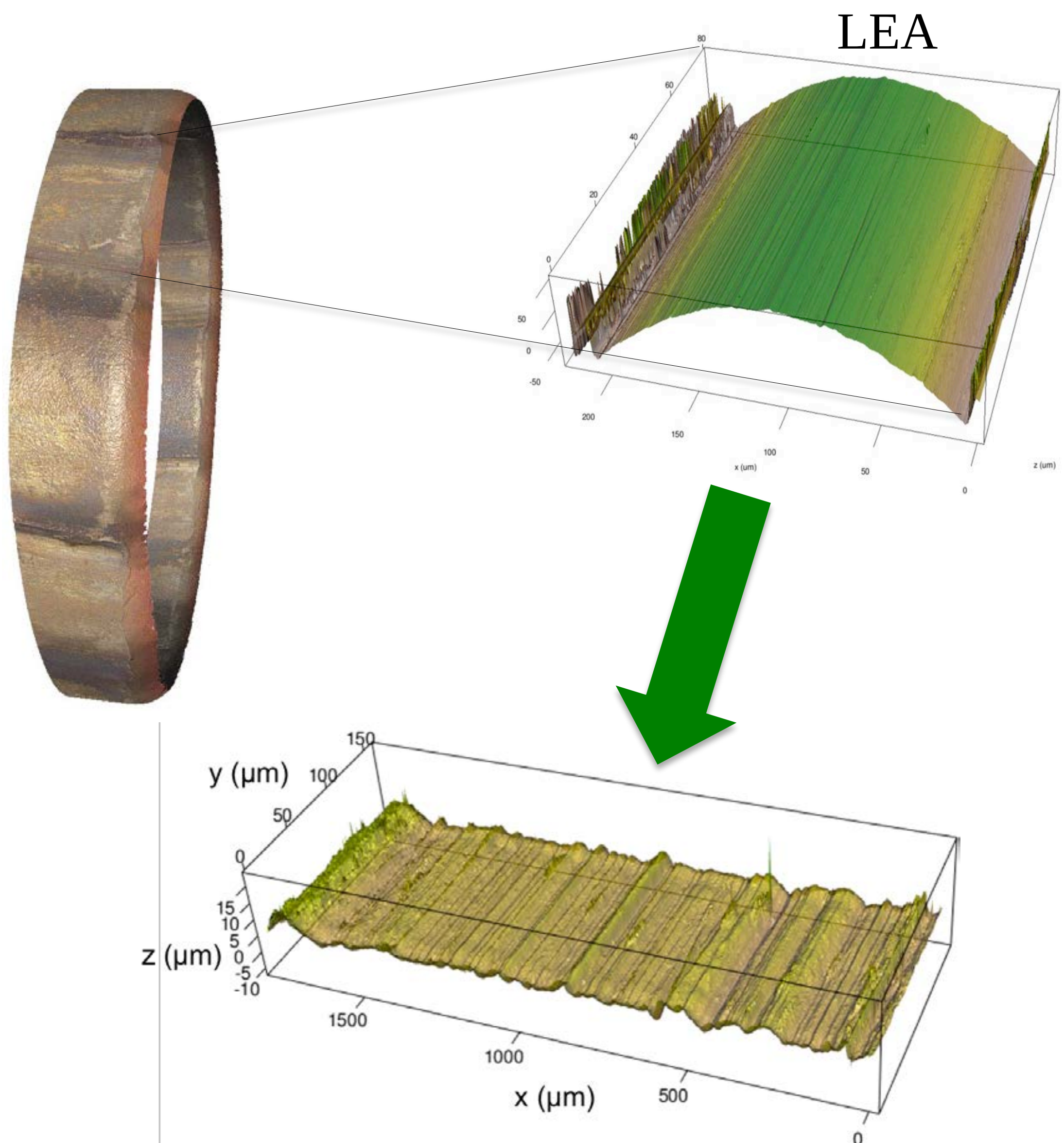
- *<u>All forms of physical evidence can be represented as numerical patterns</u>*
  - Toolmark surfaces
  - Dust and soil categories and spectra
  - Hair/Fiber categories
  - Any instrumentation data (e.g. spectra)
  - Triangulated fingerprint minutiae

- **Machine learning** trains a computer to recognize patterns
  - Can give "…the quantitative difference between an identification and non-identification"[Moran]
  - Can yield average **identification error rate estimates**
  - May be even **confidence measures for I.D.s***

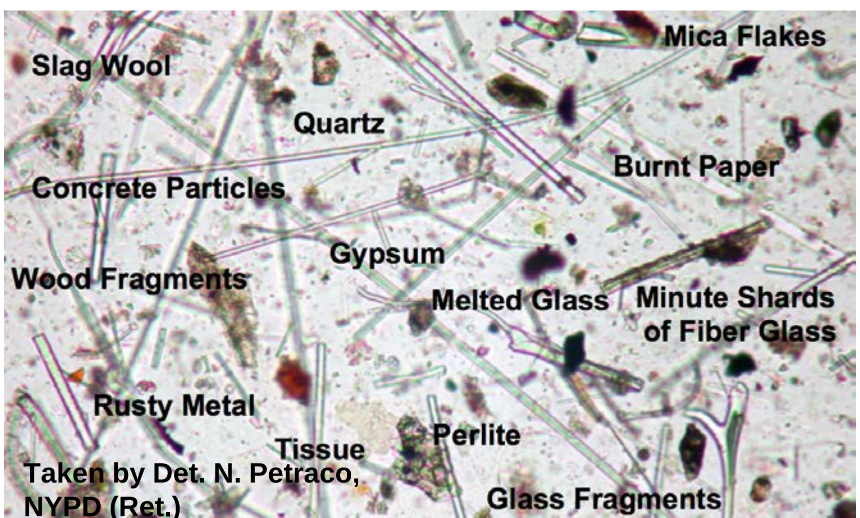
Bullets



Bullet base, 9mm Ruger Barrel



Aggregate Trace: Dust/Soils



Location Obtained____________________
Sample No.__________, Date __________, Time__________
Collector__________, Analyst__________

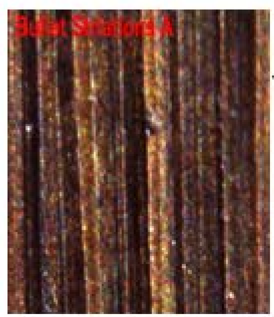
Material	Counts	Totals	%
Fiber Glass with Resin: Colorless, Red/Pink, Yellow, Blue/Black, Orange			
Rock Wool			
Asbestos			
Synthetic Fibers			
Human Hair			
Animal Hair			
Natural Fibers			
Paper Fibers			
Ceiling Tile Debris			
P. Human Remains			
Mica Flakes			
Plaster of Paris			
Concrete Debris			
Paint Smears			
Al° Fragments			
Wood Fragments			
Foam Fragments			
Charred Wood			
Charred Debris			
Perlite			
Drugs			
Plastic Fragments			
Glass Fragments			
Miscellaneous			

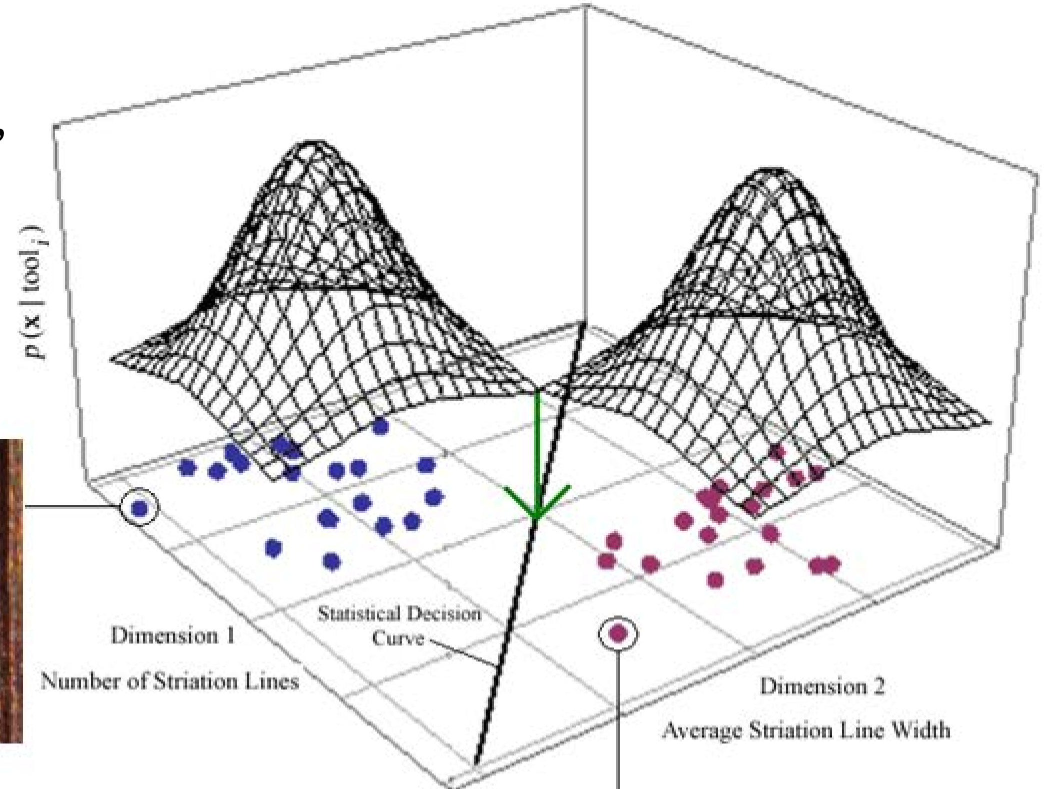
M.M. 1.539 Cargille®

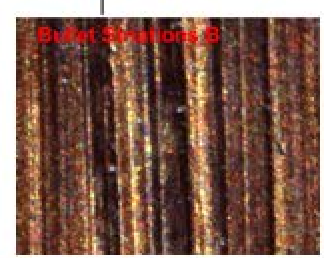
The ID-ing task

- Modern algorithms that "*make comparisons*" and "*ID unknowns*" are called **machine learning**

- Idea is to measure **features** of the physical evidence that characterize it

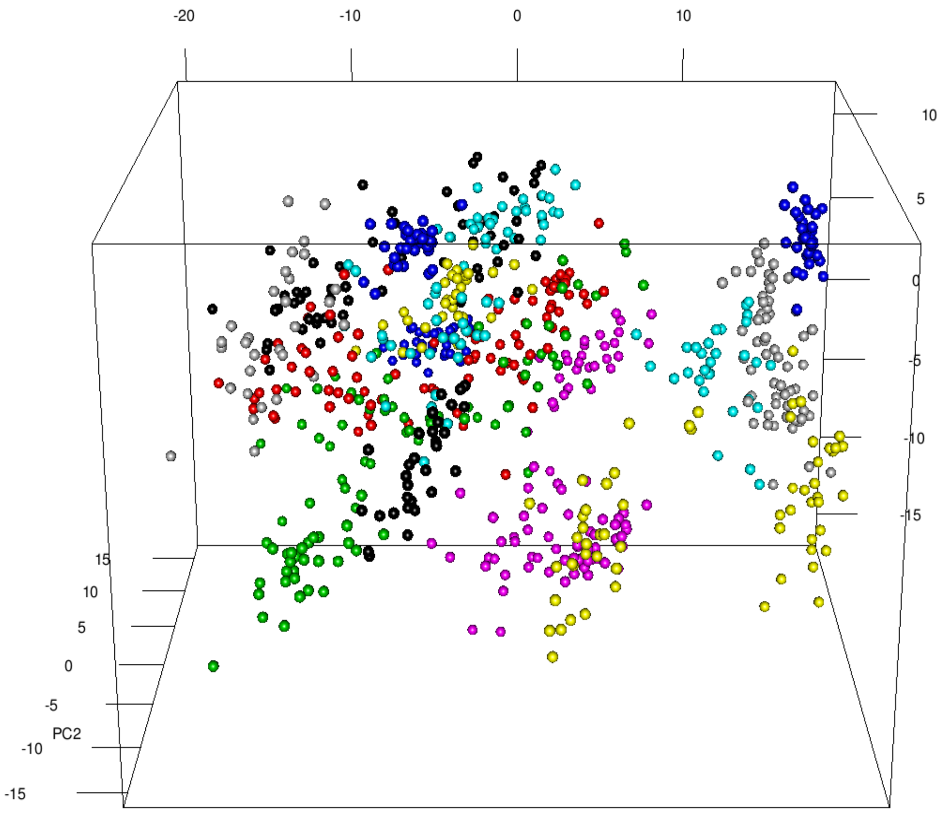




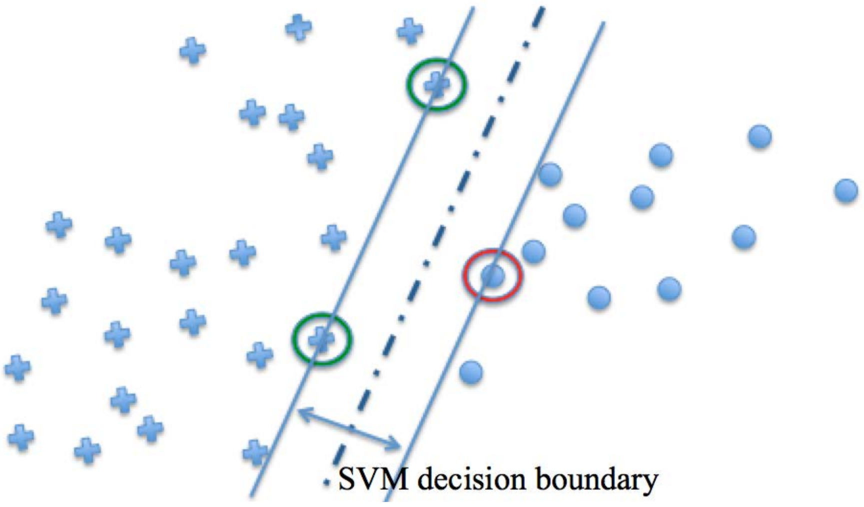


- Train algorithm to recognize "major" differences between groups of features while taking into account natural variation and measurement error.

- Visually explore: 3D PCA of 760 real and simulated mean profiles of primer shears from 24 Glocks:



- A machine learning is then trained to I.D. the toolmark:

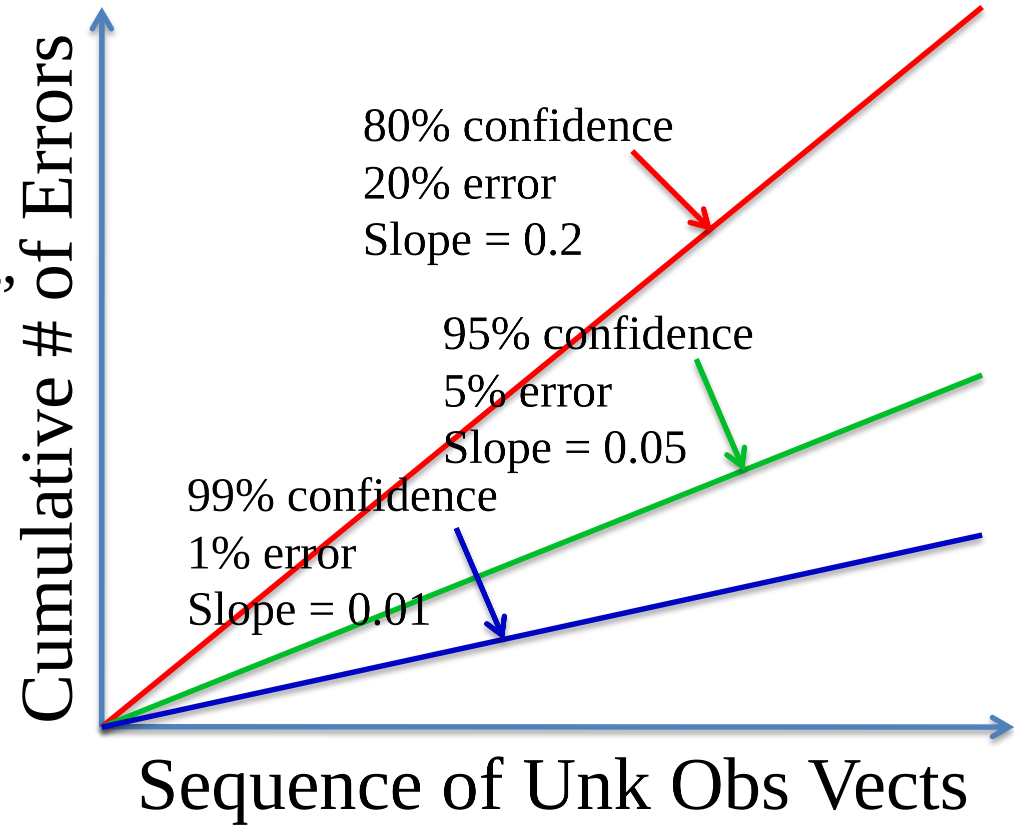


But…

How good of a "match" is it?

Conformal Prediction[Vovk]

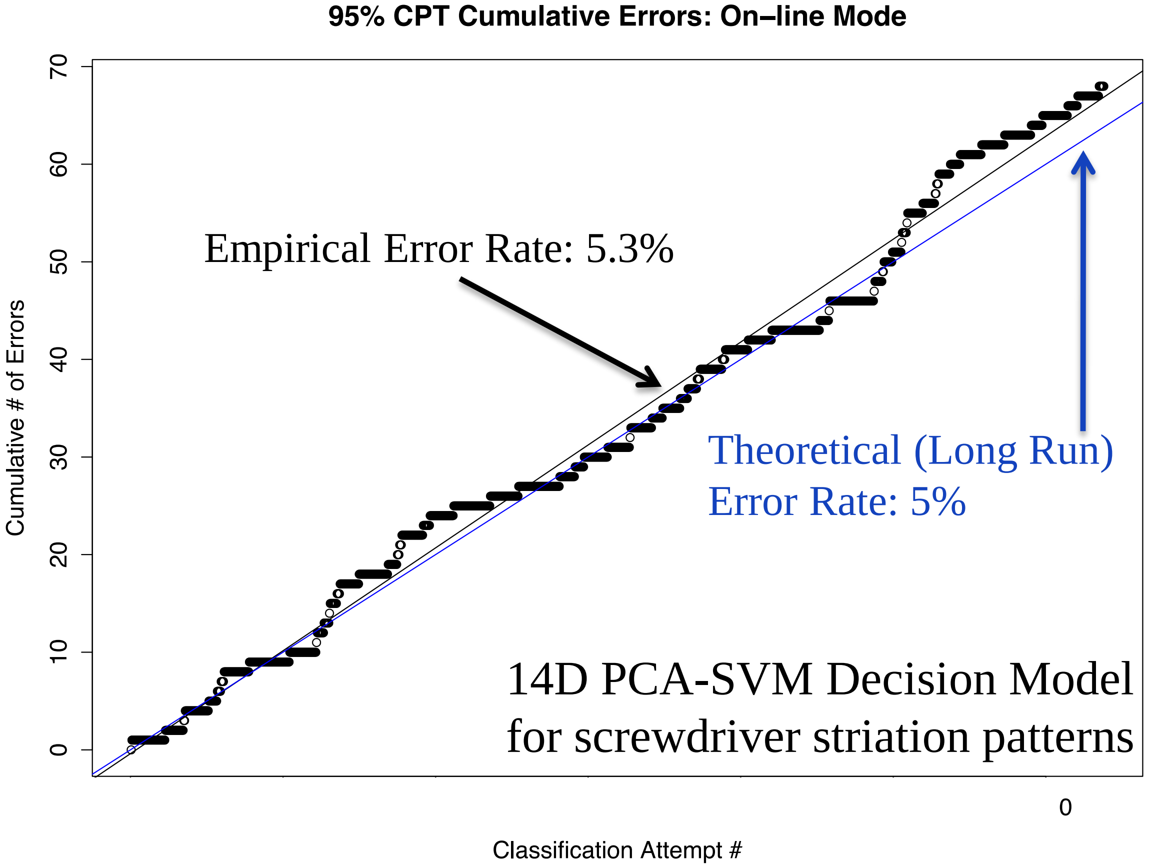
- Can give a judge or jury an easy to understand measure of reliability of classification result
  - *Confidence* on a scale of 0%-100%
  - *Testable claim*: Long run I.D. error-rate should be the chosen significance level
- This is an orthodox "frequentist" approach
  - Roots in Algorithmic Information Theory
- Data should be IID but that's it



Conformal Prediction

- For 95%-CPT (PCA-SVM) confidence intervals will not contain the correct I.D. 5% of the time *in the long run*
  - *Straight-forward validation/explanation picture for court*

cptIDPetraco for R



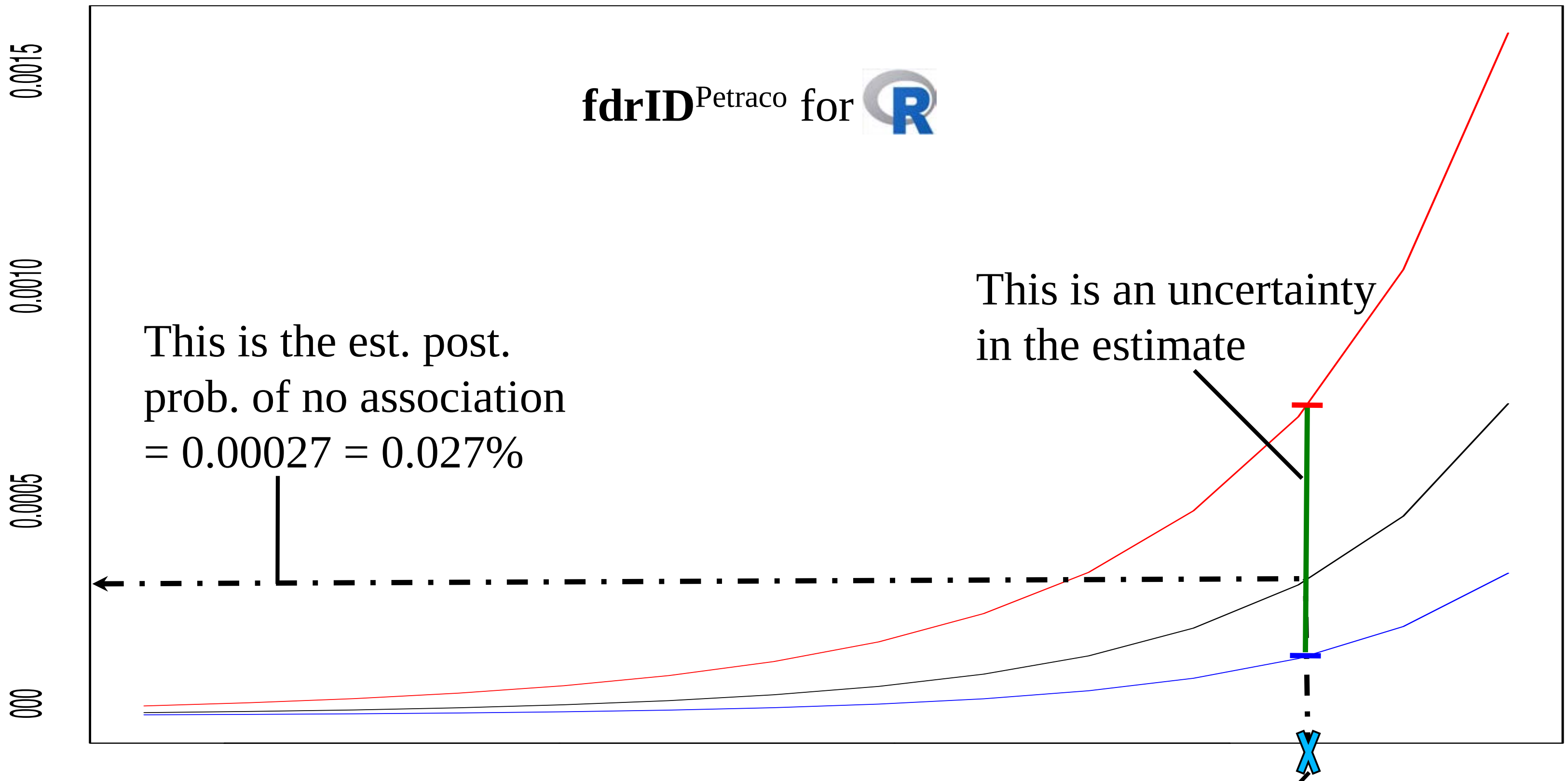
How good of a “match” is it?

Empirical Bayes

- An I.D. is output for each questioned tool mark
  - This is a computer “match”
- What’s the **probability the tool is truly the source of the tool mark**?
- Similar problem in genomics for detecting disease from microarray data
  - They use <u>*data*</u> **and** <u>*Bayes’ theorem*</u> to get an estimate

Empirical Bayes'

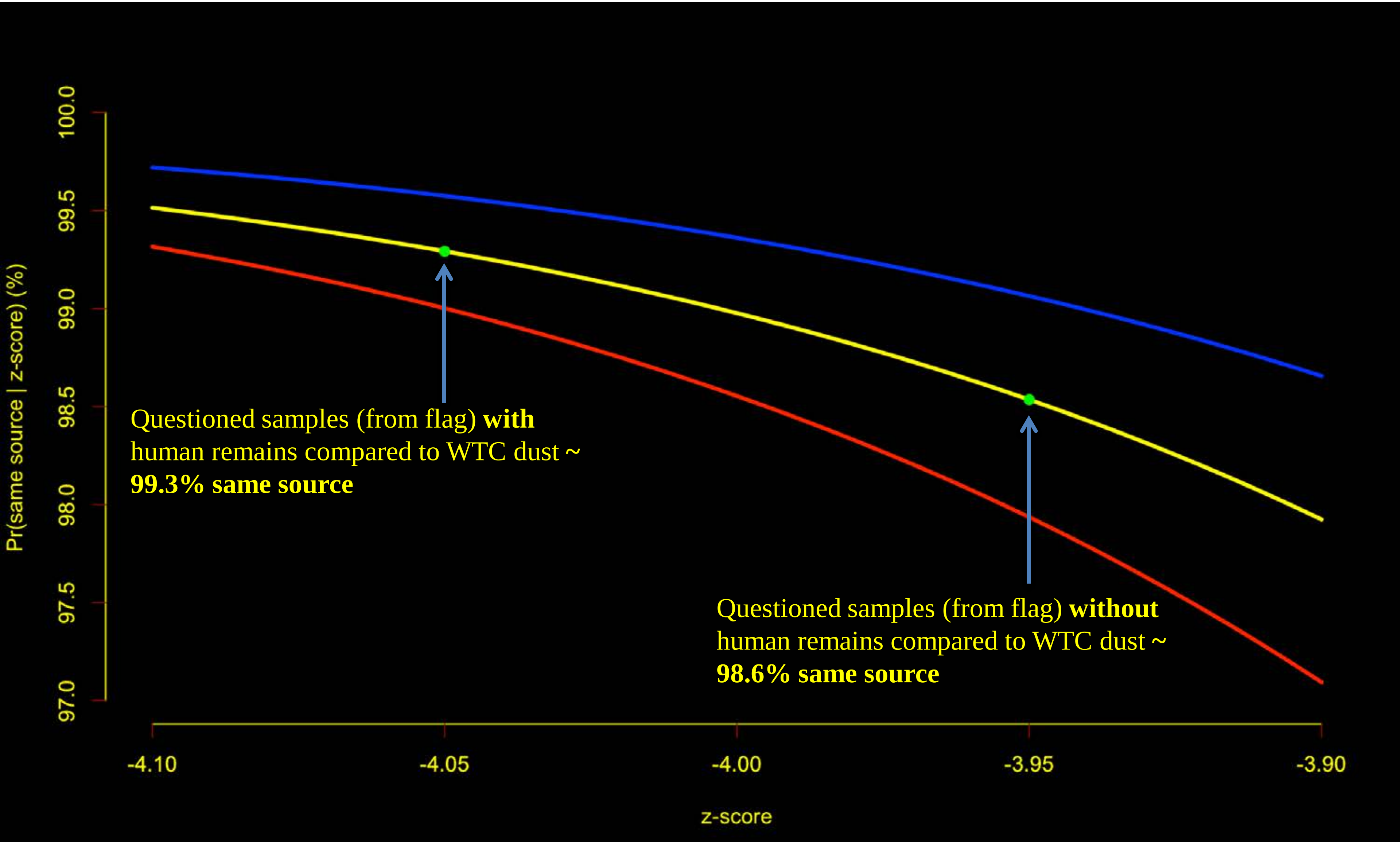
- Model's use with crime scene "unknowns":



Computer outputs "match" for:

unknown crime scene toolmarks-with knowns from "Bob the burglar" tools

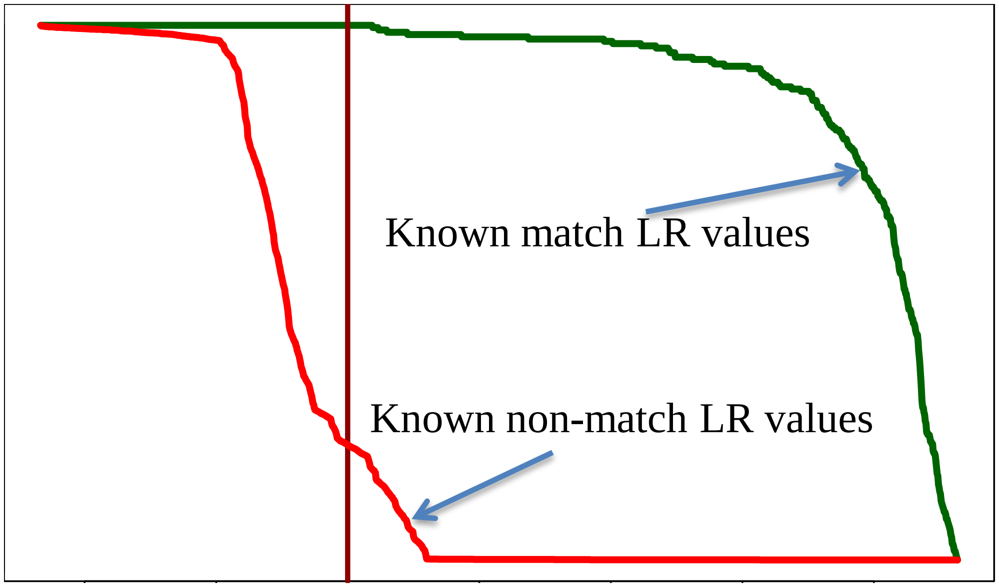
WTC Dust Source Probability (Posterior)



Likelihood Ratios from Empirical Bayes

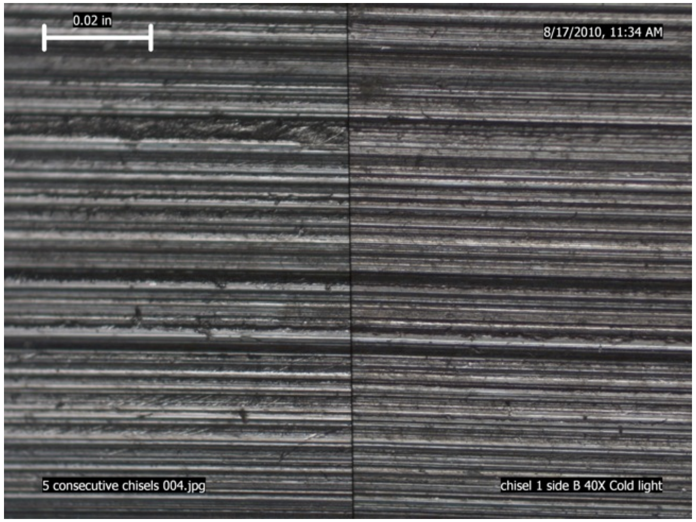
- Using the fit posteriors and priors we can obtain the likelihood ratios

$$\widehat{\mathrm{LR}}(z) = \frac{\hat{\Pr}(H_p|E)}{\hat{\Pr}(H_d|E)} \Bigg/ \frac{\hat{\Pr}(H_p)}{\hat{\Pr}(H_d)} = \frac{\widehat{\mathrm{tdr}}(z)}{\widehat{\mathrm{fdr}}(z)} \frac{\hat{\pi}_0}{1-\hat{\pi}_0}$$



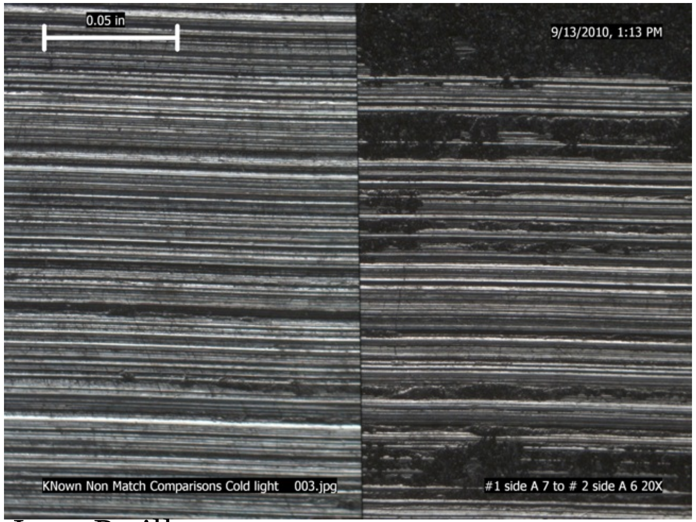
Typical "Data Acquisition" For Toolmarks

KM



Jerry Petillo

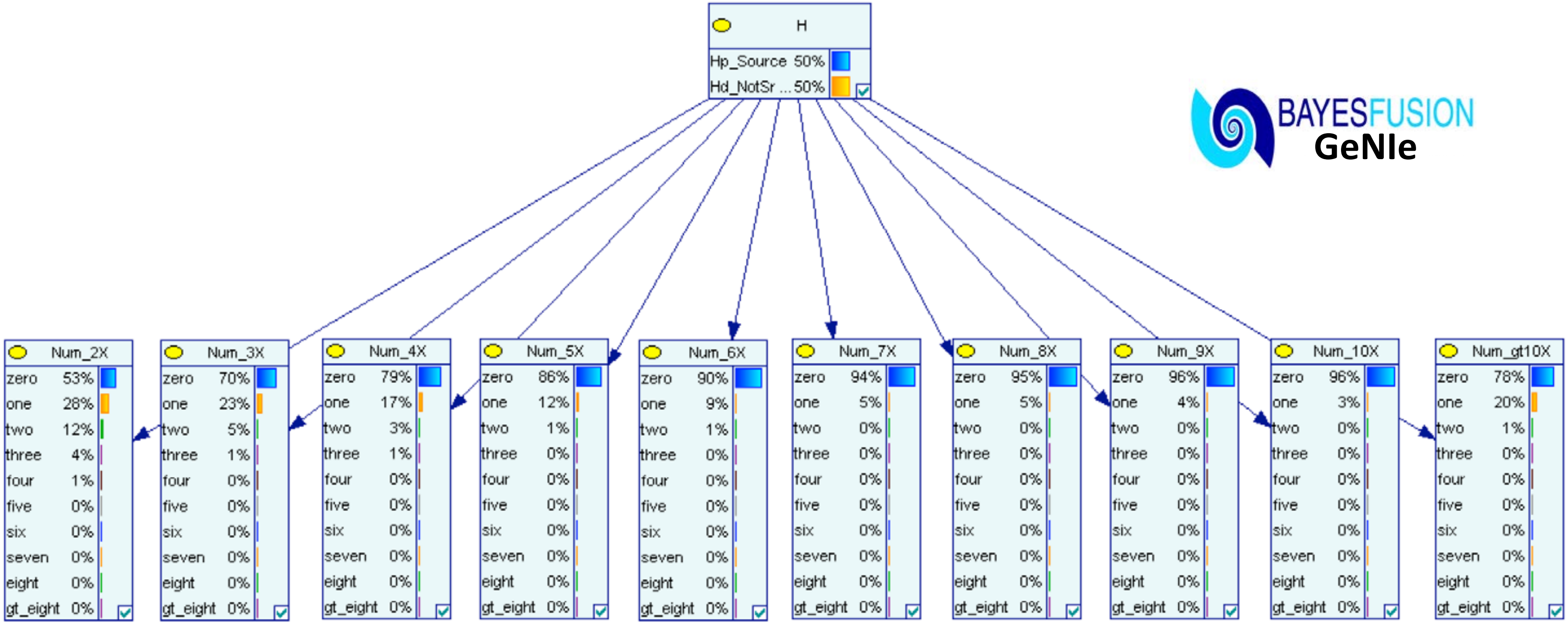
KNM



Jerry Petillo

Bayesian Networks

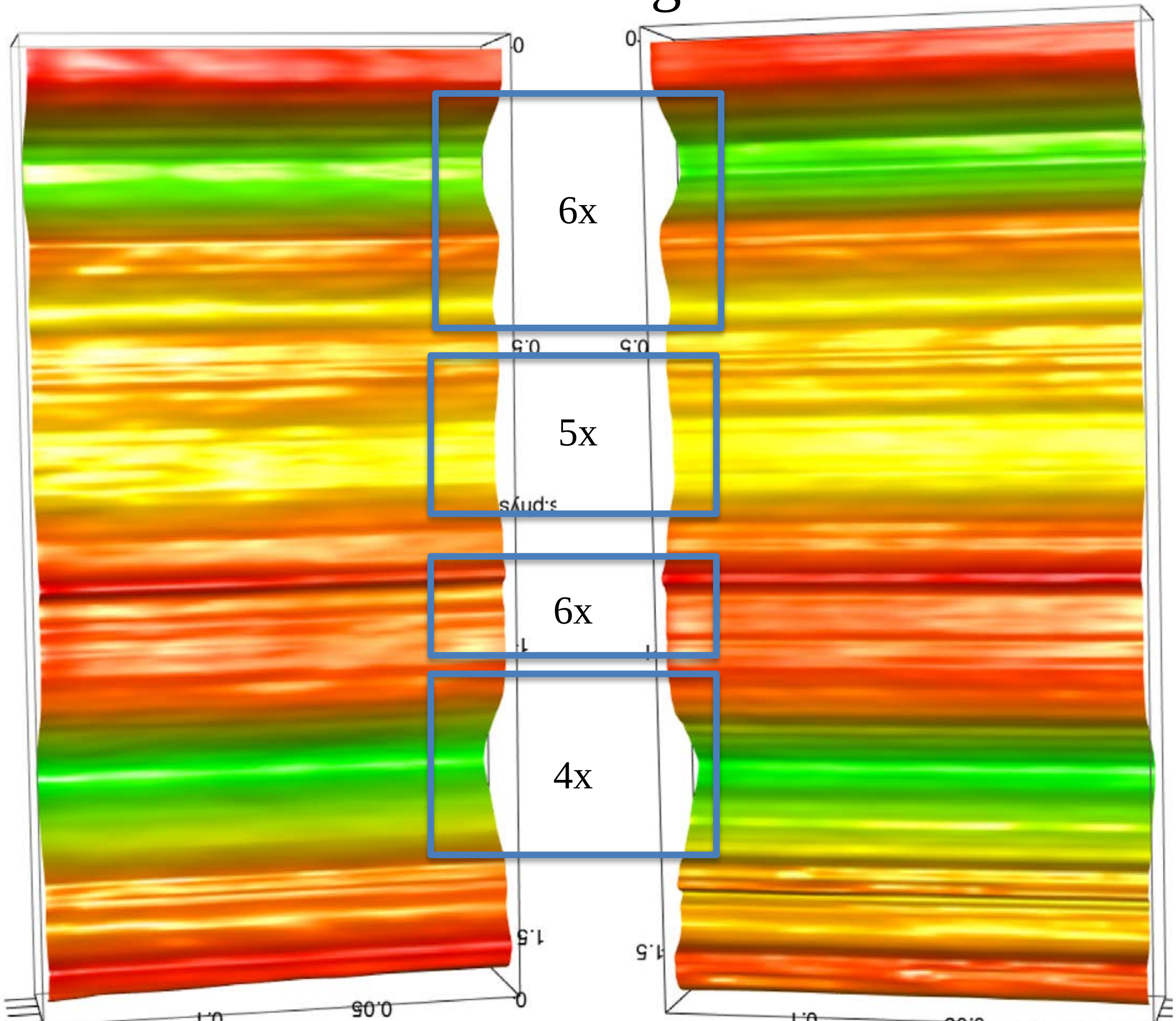
“Prior” network based on historical/available count data and multinomial-Dirichlet model for run length probabilities:



Run Your "Algorithm"

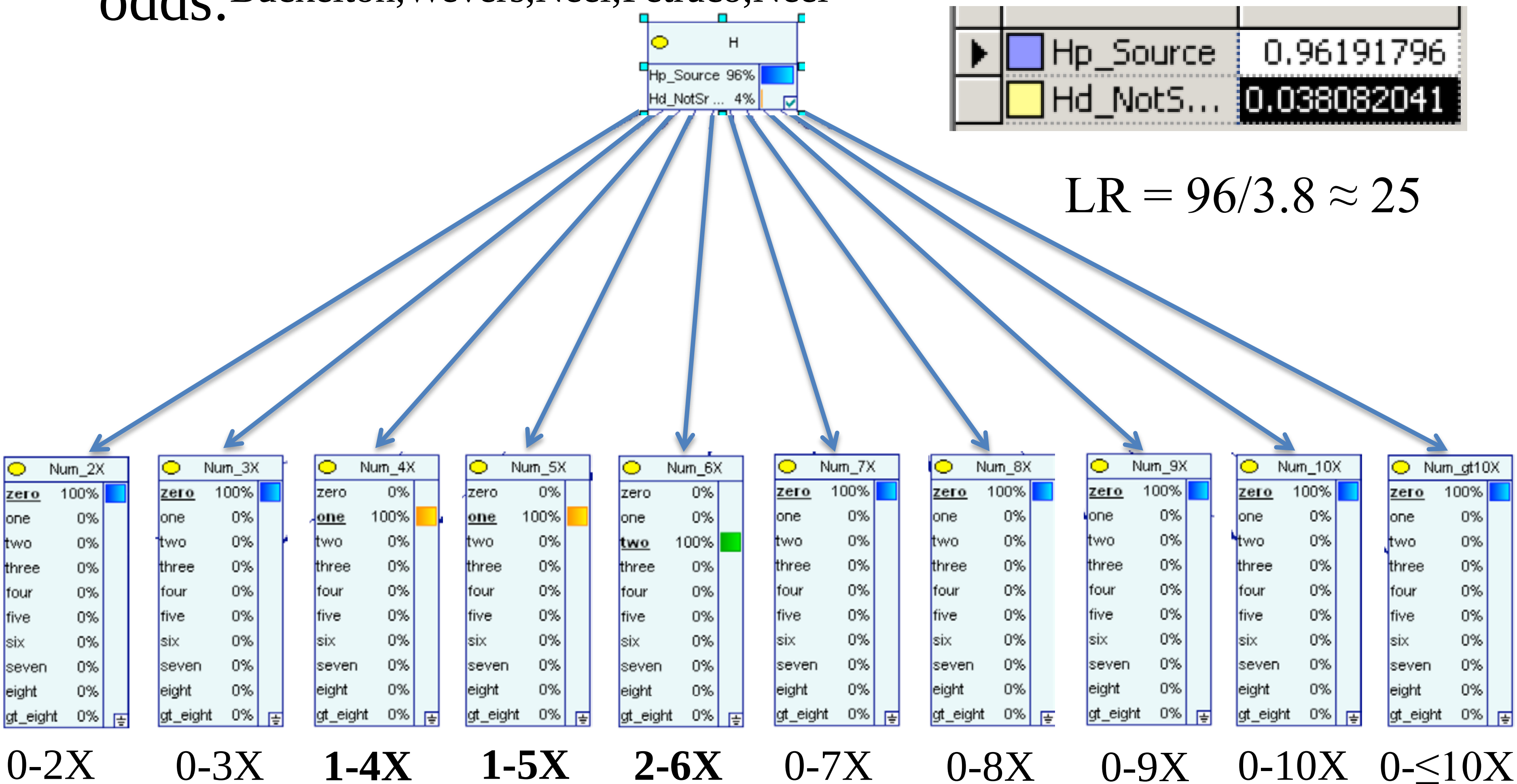
Known

Unknown



1-4X, 1-5X, 2-6X

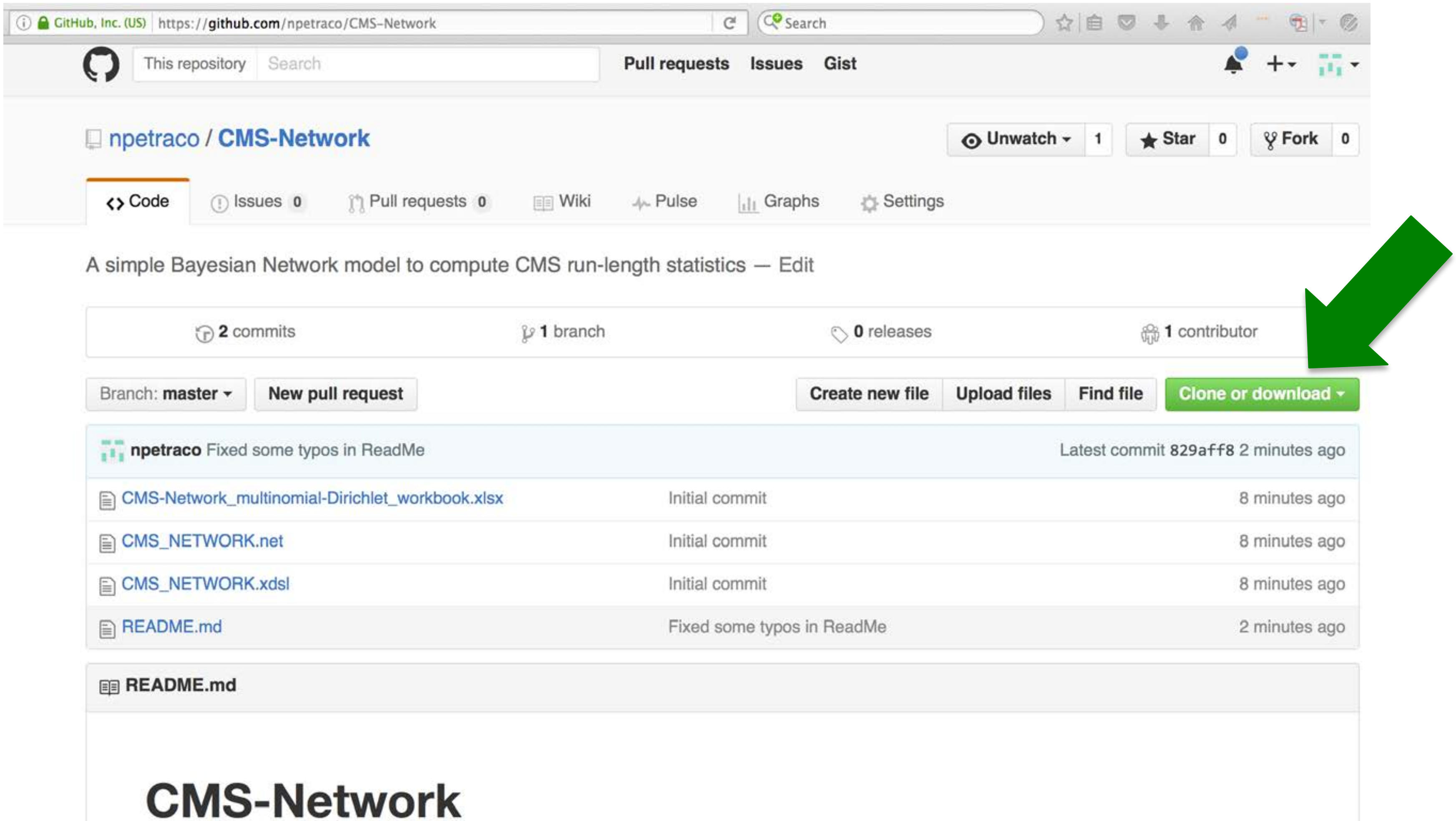
Enter the observed run length data for the comparison into the network and update “match” (same source) odds:Buckelton,Wevers,Neel;Petraco,Neel



LR = 96/3.8 ≈ 25

The evidence “strongly supports”Kass-Raftery that the striation patterns were made by the same tool

Where to Get the Model and Software

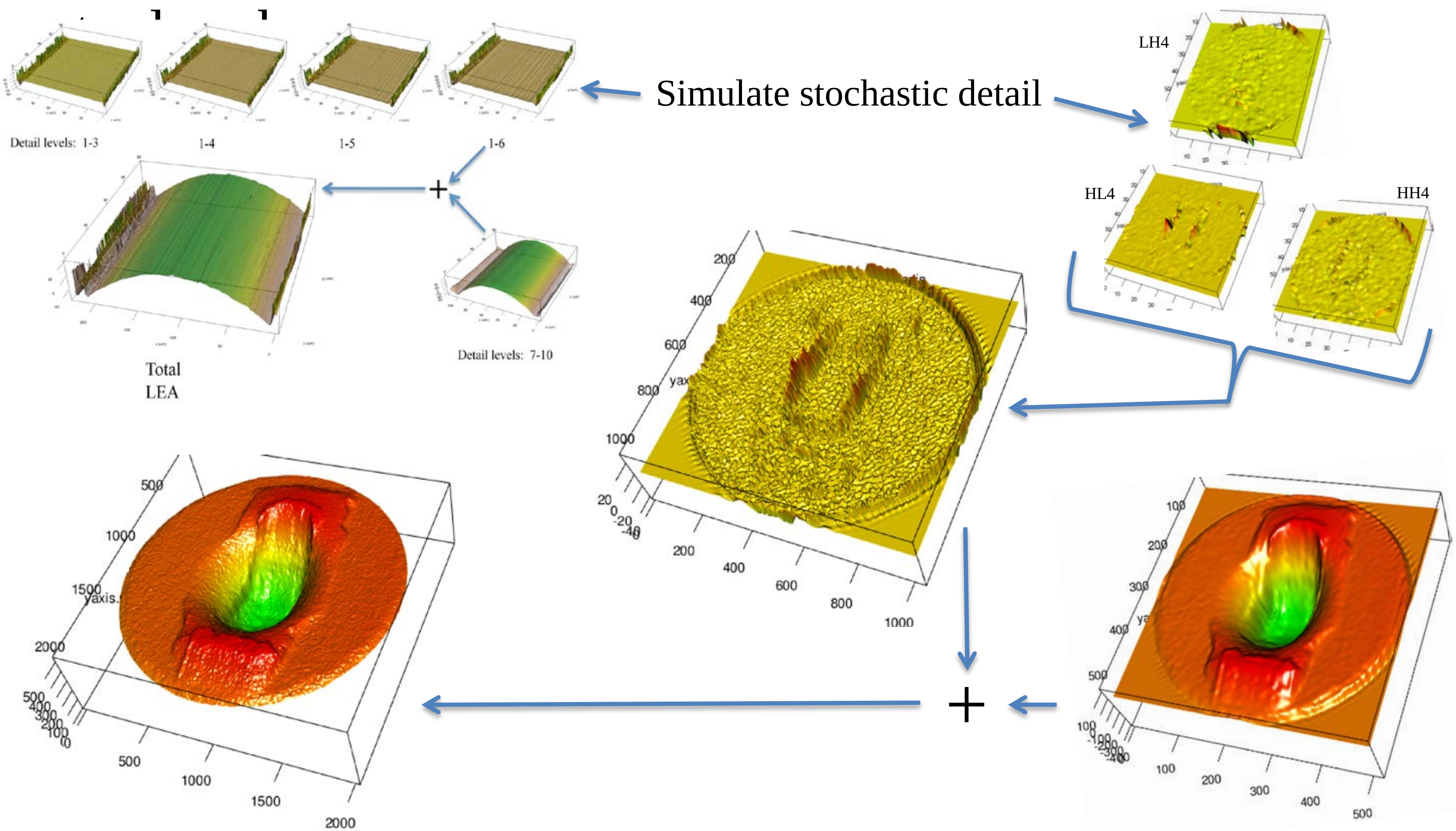


Bayes Net software: No cost for non-commercial/demo use

BayesFusion: http://www.bayesfusion.com/
SamIam: http://reasoning.cs.ucla.edu/samiam/
Hugin: http://www.hugin.com/
gR packages: http://people.math.aau.dk/~sorenh/software/gR/

Directions for the future

- **Data!** Working on a wavelet based simulator for 2D

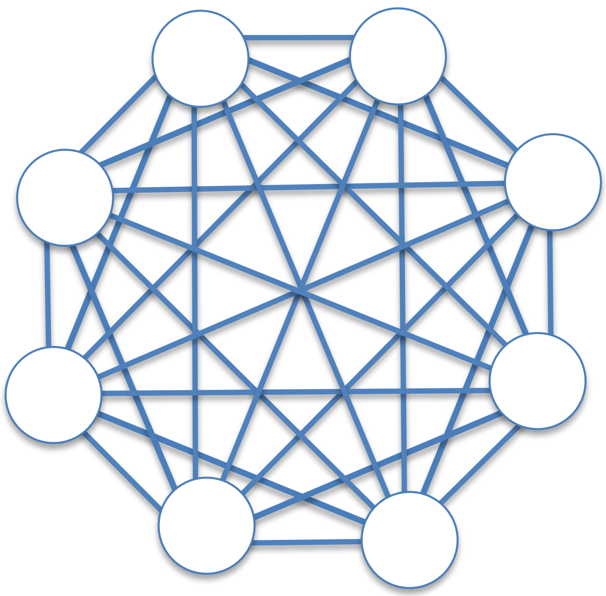


Data!: Aggregate evidence dependencies

Dust Tabulation Sheet   Dust Specimen No.____________   Date Acquired____________

Dust specimen source: ____

Human Hair	EUA Brn	EUA Blk	EUA Gry	EUA Red	EUA Blonde	AS	AFA	India	MIXED		
Head Natural											
Head Treated											
Pubic											
Other Body Area											
Animal Hair	White	Brown	Beige	Black	Gray	Red	W/Brn	W/Blk	W/Beige	W/Gray	Banded
Dog											
Cat											
Rabbit											
Mink											
Rodent											
Other ( )											
Synthetic Fibers	Red	Blue	Green	Orange	Brown	Black	Violet	Pink	Yellow	Other	Colorless
Rug Olefin Triangular X-S											
Rug Nylon "											
Rug Polyester "											
Olefin Any X-S											
Nylon "											
Polyester "											
Acrylic Dog-bone X-S											
Acrylic Round X-S											
Acrylic Other X-S											
Modacrylic											
Rayon											
Acetate/ Triacetate											
Other ( )											
Natural Fibers	Red	Blue	Green	Orange	Brown	Black	Violet	Pink	Yellow	Other	Colorless
Wool											
Silk											
Cotton											
Hemp											
Jute											
Sisal											
Manila											
Feathers ( )											
Other ( )											
Mineral Fibers	Pink	Red	Clear	Other							
Fiber Glass Fibers (Resin)											
Slag/Mineral Wool											
Asbestos											
Glass/Mineral Grains	Clear	Green	Amber	Brown	Blue	Other					
Glass Chips											
Safety Glass (Dice Pieces)											
Quartz											
Calcite											
Feldspar											
Micas											
Other ( )											
Miscellaneous	White	Red	Blue	Green	Orange	Brown	Black	Violet	Yellow	Other	Colorless
Paint Chips											
Glitter (Color/Shape)											
Plaster Chips/Grains											
Concrete Dust											
Brick Dust											
Saw Dust											
Vegetable Matter	Seeds	Leaves	Twigs	Bark							
Starch Grains	Potato	Rice	Corn	A-Root							
Salt Grains											
Sugar Granules											
Insect Parts	Roach	Ants	Bees	Moths	Other						
Other ( )											



- Ising and Potts-like “spin” models:
  - Capture dependences between components of materials/trace evidence
  - Simulate from model using standard Stat-Mech techniques

General Weight of Evidence Evaluations:

- **Weight of evidence** IMHO was canonically defined by Jefferys:

"Weight of evidence" = **Bayes Factor**

"**Thermodynamic Integration**" trick:

$$\ln\left(p(D|H_i)\right) = \int_0^1 \frac{d\ln(Z_\beta)}{d\beta}\, d\beta \qquad \frac{\Pr(H_p|D)}{\Pr(H_d|D)} = \frac{p(D|H_p)}{p(D|H_d)}\frac{\Pr(H_p)}{\Pr(H_d)}$$

$$\frac{d\ln(Z_\beta)}{d\beta} = \mathrm{E}_{\boldsymbol{\theta}_i}\left[\ln\left(p(D|\boldsymbol{\theta}_i)\right)\Big|D,\beta\right]$$

We can (in principle…) get these averages from MCMC for fixed β from 0 to 1 and numerically integrate the above[Lartillot,Friel,Gelman]

More Future Directions

- **<u>Clean up:</u>** cptID, fdrID
- **<u>GUI modules</u>** for common toolmark comparison tasks/calculations using 3D microscope data
- **<u>2D features</u>** for toolmark impressions: **feature2**[Petraco]
- **<u>Parallel/GPU/FPGA</u>** implementation of computationally intensive routines e.g. ALMA Correlator for astronomy data
  - Especially for retrieving "relevant pop/best match" reference sets
- **<u>Uncertainty for Bayesian Networks</u>**
  - Models, parameters…

References

Petraco:

- https://github.com/npetraco/x3pr
- https://github.com/npetraco/feature2
- https://github.com/npetraco/cptID
- https://github.com/npetraco/fdrID

Whitcher: https://cran.r-project.org/web/packages/waveslim/index.html

R-Core: https://www.r-project.org/

Vovk: http://www.alrw.net/

Efron:

- Efron, B. "Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction", Cambridge, 2013.

Neel:

- Neel, M and Wells M. "A Comprehensive Analysis of Striated Toolmark Examinations. Part 1: Comparing Known Matches to Known Non-Matches", AFTE J 39(3):176-198 2007.

Wevers:

- Wevers, G, Michael Neel, M and Buckleton, J. "A Comprehensive Statistical Analysis of Striated Tool Mark Examinations Part 2: Comparing Known Matches and Known Non-Matches using Likelihood Ratios", AFTE J 43(2):1-9 2011.

Buckleton:

- Buckleton J, Nichols R, Triggs C and Wevers G. "An Exploratory Bayesian Model for Firearm and Tool Mark Interpretation", AFTE J 37(4):352-359 2005.

BayesFusion: http://www.bayesfusion.com/

Acknowledgements

- Professor Chris Saunders (SDSU)
- Professor Christophe Champod (Lausanne)
- Alan Zheng (NIST)
- Ryan Lillien (Cadre)
- Scott Chumbley (Iowa State)
- Robert Thompson (NIST)
- Research Team:
    - Mr. Daniel Azevedo
    - Dr. Martin Baiker
    - Dr. Peter Diaczuk
    - Mr. Antonio Del Valle
    - Ms. Carol Gambino
    - Dr. James Hamby
    - Ms. Lily Lin
    - Mr. Kevin Moses
    - Mr. Mike Neel
    - Ms. Alison Hartwell, Esq.
    - Dr. Brooke Kammrath
    - Mr. Chris Lucky
    - Off. Patrick McLaughlin
    - Dr. Mecki Prinz
    - Dr. Linton Mohammed
    - Mr. Nicholas Petraco
    - Ms. Stephanie Pollut
    - Dr. Jacqueline Speir
    - Dr. Peter Shenkin
    - Mr. Peter Tytell
    - Dr. Peter Zoon

Collaborations,
Reprints/Preprints:
npetraco@gmail.com
http://jjcweb.jjay.cuny.edu/npetraco/