SAND2011-9039 C

Exceptional service in the national interest

SAND2011-9039C

Sandia National Laboratories

Photos placed in horizontal position
with even amount of white space
between photos and header

A Mathematician turned Statistician: From Academia to a National Laboratory

Stephanie Fitchett
sfitche@sandia.gov

8 December 2011

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Outline

- My background
- Sandia National Laboratories and the statistics group at Sandia
- Some projects I have worked on since joining Sandia
  - Estimating multivariate probability distributions
  - Shock spectra (colleague's project)
  - Aircraft inspection
  - Mini-SAR imagery
- Non-Sandia projects in epidemiology
- Internship opportunities and more information

My Background

- Pure Mathematics
  - B.S., M.S., and Ph.D. in mathematics from the University of Nebraska
  - Teaching postdoc
  - Florida Atlantic University's Honors College
- Statistics
  - M.S. in Statistics from Colorado State University
  - Work with USDA
- University of Northern Colorado
- Joined Sandia in late January 2011

Sandia

- Develops science-based technologies that support our national security
- Employs around 8500 full-time, another 1000 contractors, and about 500 students, postdocs, other limited-term positions
  - Lots of engineers of all flavors
- Environment that supports professional development and education
  - Opportunities to change jobs in the lab, training opportunities
- Good place to work

The Statistics Group

- 6 statisticians that act as consultants to projects across the lab
  - 4 with Ph.D. in statistics, 1 with Ph.D. in math, 1 with M.S. in math
- Huge variety of work, large and small projects
  - Reliability
  - Quantification of Margins and Uncertainty
  - Teaching
  - Design of experiments
  - Sampling and monitoring plans
  - Data analysis

A SAMPLE OF PROJECTS

Estimating probability distributions

- Measurements on ~ 30 variables are collected for different parts
  - Variables are discrete, with perhaps 10-20 possible values for each
- Want to be able to identify the part based on the measurements
- First step: understand the joint distribution of the measurements (10^{30} possible combinations of values) by using an approximation to the distribution
- Chow-Liu algorithm
  - Compute "mutual information" for each pair of variables
  - Use Kruskal's algorithm to find maximum spanning tree
  - Approximate the desired distribution by creating the distribution determined by the spanning tree
- Use simulations to decide if the algorithm provides a sufficient approximation

Small example:
Data, marginals, two-ways

X_1	X_2	X_3	X_4	X_5
5	23	17	9	16
4	22	18	8	16
5	26	18	8	16
1	25	16	9	16
1	22	18	8	15
1	26	17	9	16
3	22	19	7	15
5	24	18	8	16
…	…	…	…	…

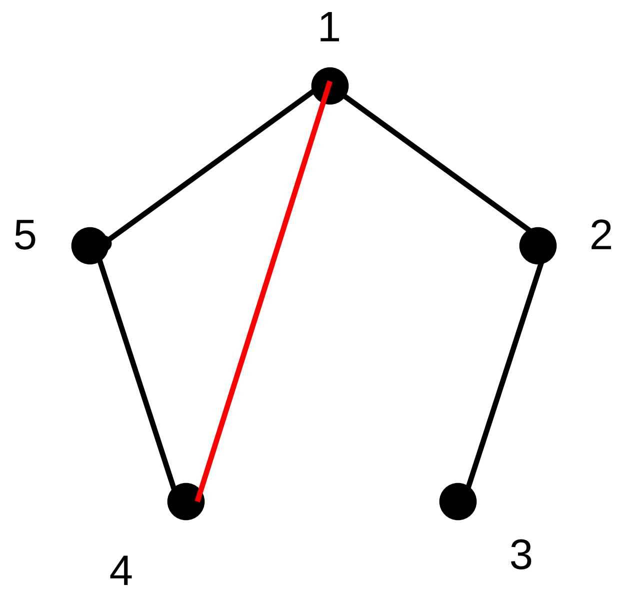
	22	23	24	25	26	
1		0.05	0.08	0.04	0.01	0.18
2	0.15	0.04	0.06			0.25
3	0.11	0.01	0.03	0.10	0.01	0.26
4	0.03	0.02	0.03	0.04	0.03	0.15
5	0.08			0.08		0.16
	0.37	0.12	0.20	0.26	0.05	

Full Joint Distribution could be represented as a 5-dimensional array (5 x 5 x 4 x 3 x 2, in this case)

Example with 5 variables

$$\hat{I}(X_i, X_j) = \sum_{x_i, x_j} P(x_i, x_j) \log_2 \frac{P(x_i, x_j)}{P(x_i)P(x_j)}$$

Pair	I-hat
23	0.638
15	0.427
45	0.349
14	0.262
12	0.082
35	0.031
13	0.028
34	0.019
24	0.012
25	0.009

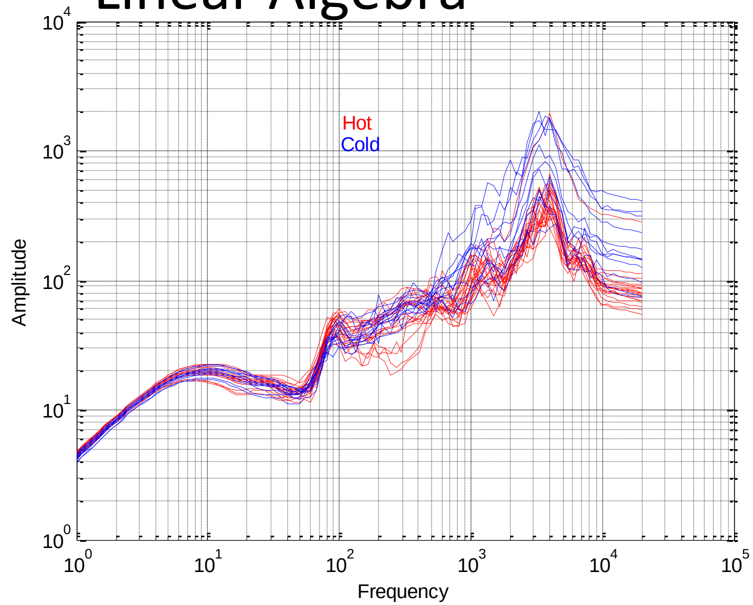


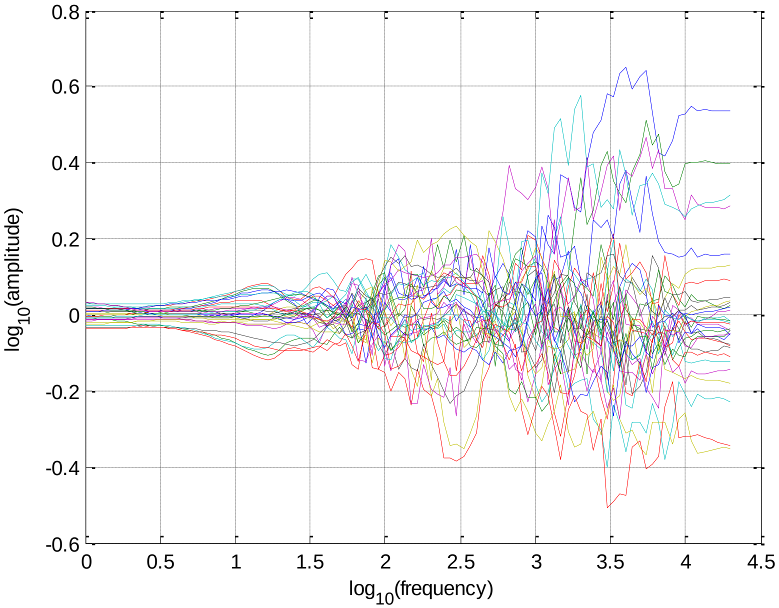
$$P_{Chow\text{-}Liu}(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_5|x_1)P(x_4|x_5)$$

$$= P(x_1)P(x_2,x_1)P(x_3,x_2)P(x_5,x_1)P(x_4,x_5)/[P(x_1)^2P(x_2)P(x_5)]$$

Shock spectra (colleague's project)

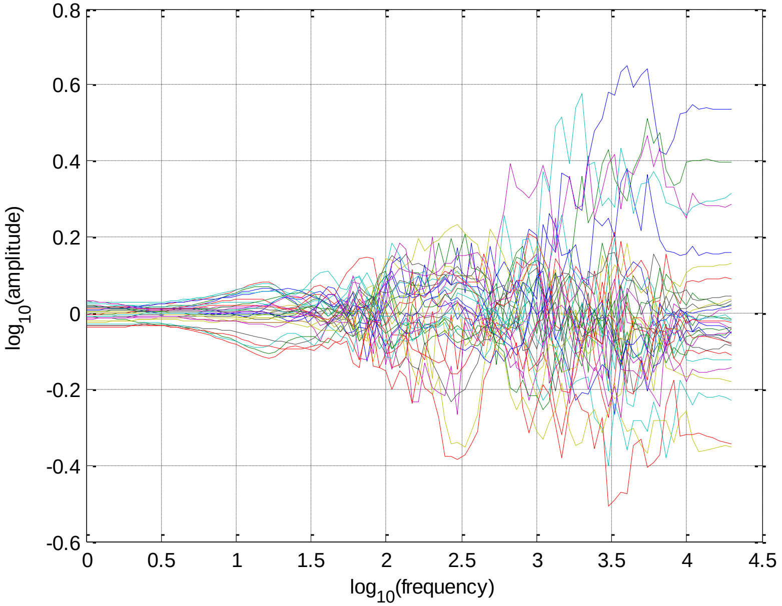
- Given some shock spectra, estimate largest probable shock that the ejected object could see
- For each frequency, find the mean amplitude across all spectra and subtract the mean amplitude from each spectrum
- Create a matrix of the mean-centered spectra and use lots of Linear Algebra





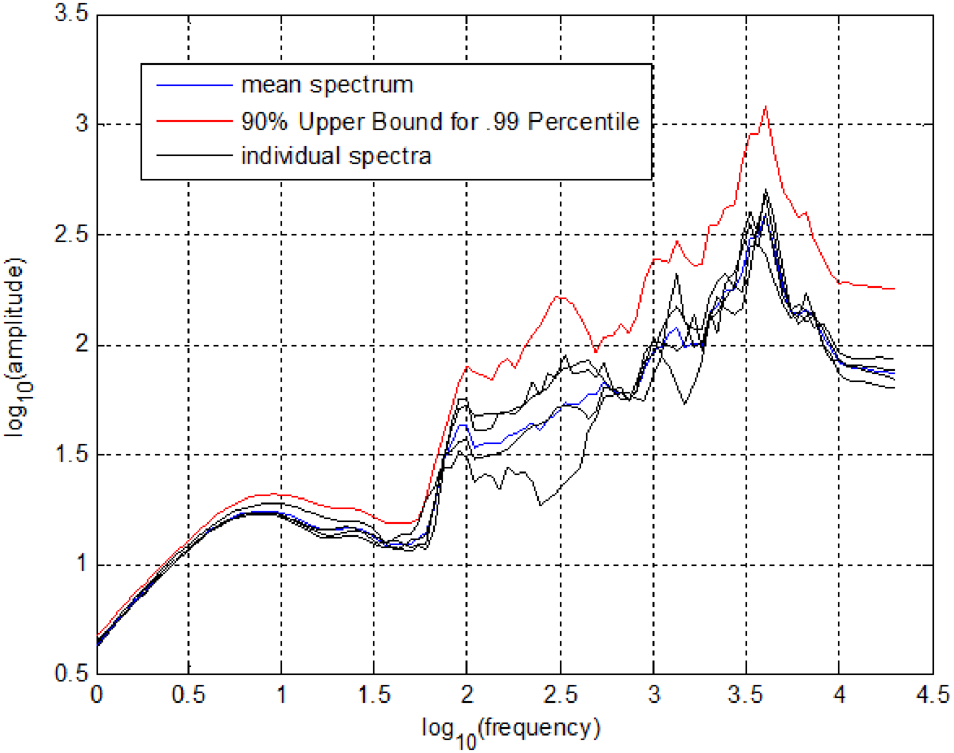
Shock spectra

- Create matrix X with one row for each (mean-centered) spectrum, one column for each frequency
- Find an eigenvalue decomposition of XX^t
  - $XX^t = Q\Lambda Q^t$ where the columns of Q are the eigenvectors of XX^t and Λ is a diagonal matrix with the eigenvalues on the diagonal
- Rearrange Q and Λ so that eigenvalues are in decreasing order
- Reduce to a few columns of Q and Λ



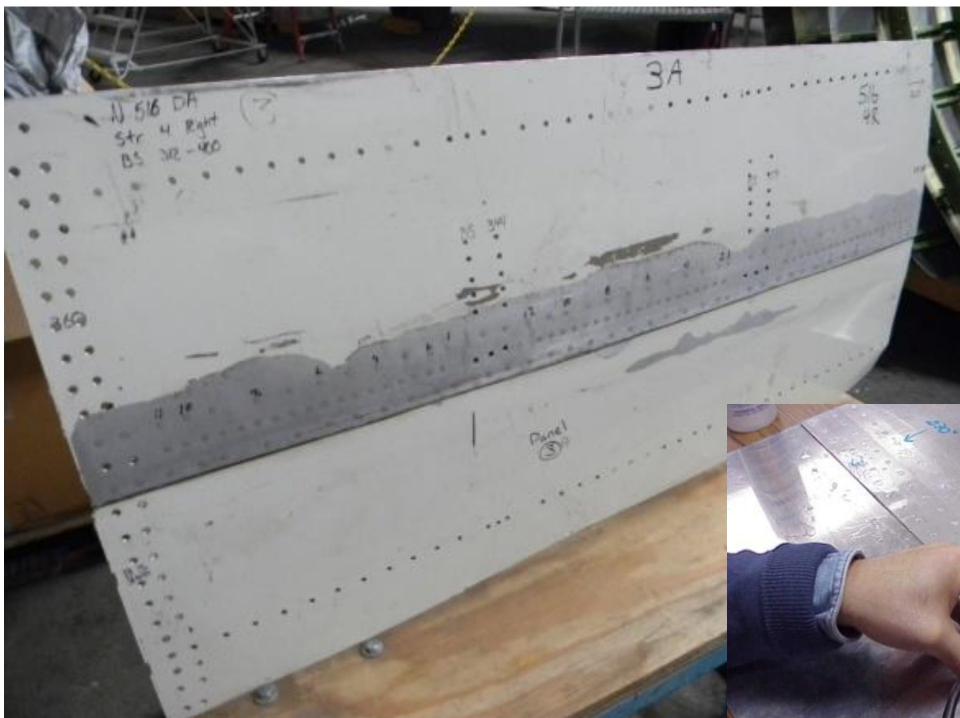
Bounding the spectra

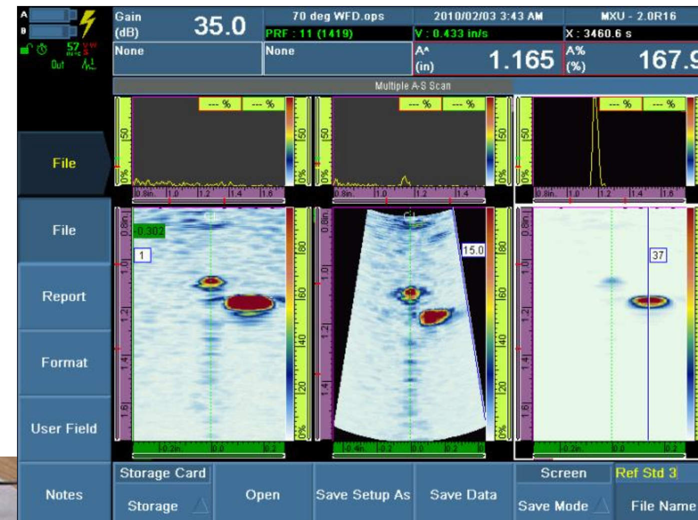
- Estimate spectra using orthogonal “scores” (the principal eigenvectors)
- Estimate the mean spectrum and standard deviations for each frequency
- Assuming spectra are normally distributed at each frequency, we can estimate 99th percentile (and bootstrap an upper bound)
- All relies heavily on the <u>orthogonality</u> of the eigenvectors



Ensuring the safety of aircraft

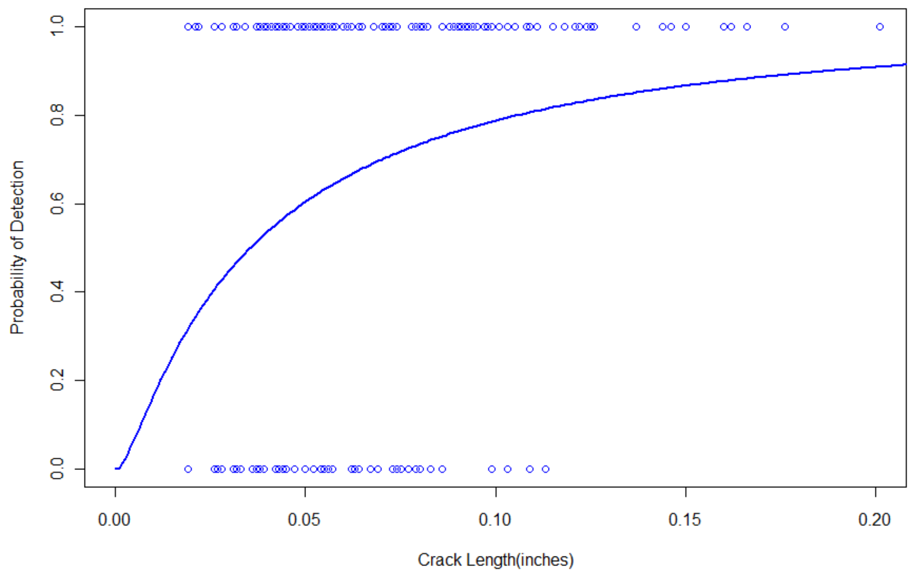
- Sandia studies non-destructive inspection techniques
  - look for cracking in aluminum aircraft skins





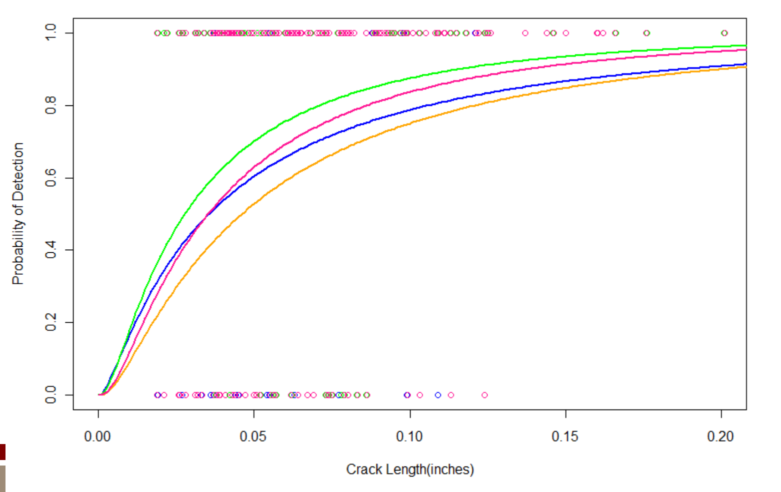
Evaluating an inspection technique

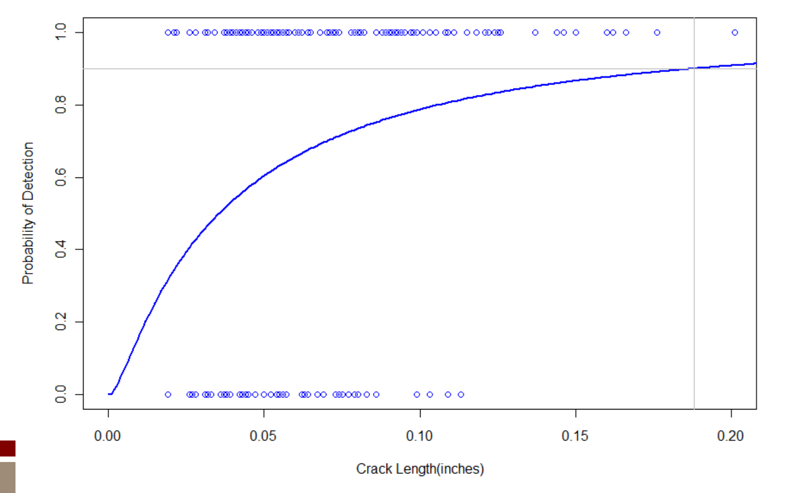
- Build specimen sets with known defects
- Have inspectors inspect the sets
  - Record information on signal return and detection of flaws at each site (bolt hole)
- Fit a probability of detection curve
- Compare inspection methods



Things to consider

- Difference in operators
- How realistic is the specimen set to real aircraft?
  - Can we get good information from fabricated specimens?
- Once we have a probability of detection curve for an inspection method, how can it inform the inspection schedule to ensure safety?
- New composite aircraft



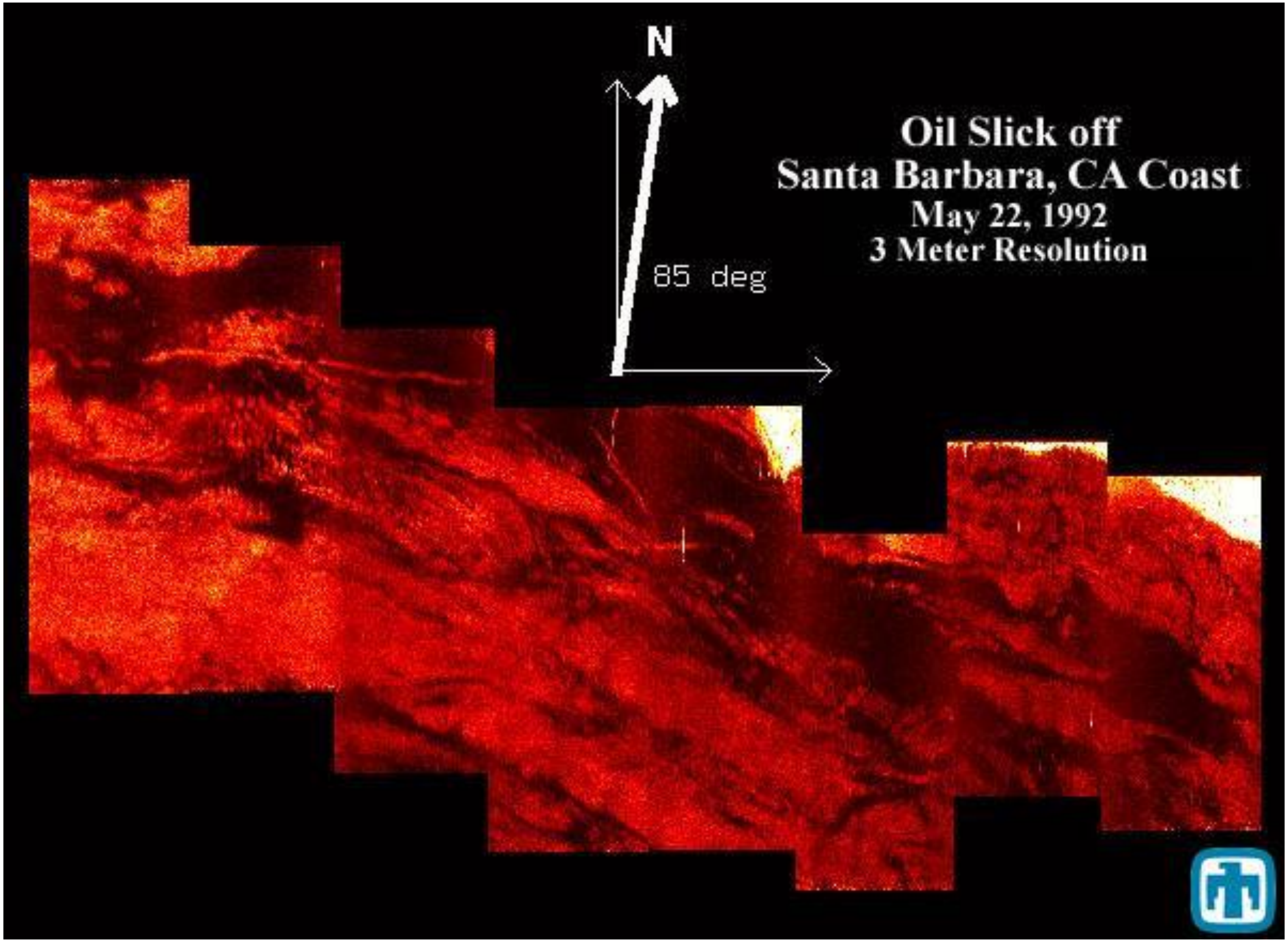


Mini-SAR imagery

SAR

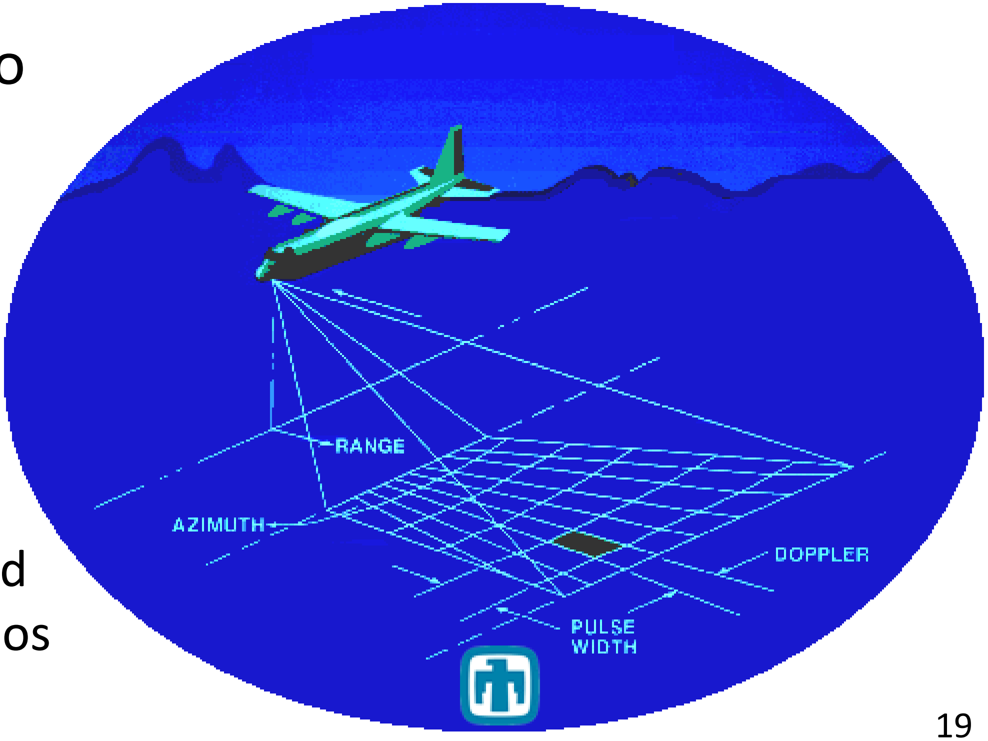
- SAR = Synthetic Aperture Radar
- Work in all weather conditions, day or night
- Distinguish terrain features because different materials return signal differently
- Recognize and identify man-made objects
- Environmental applications
  - Monitoring crop characteristics, deforestation, ice flows, oil spills
- At low frequencies, ability to see through foliage and soil
  - Identify buried utility lines, for example
- Public website: www.sandia.gov/RADAR/

Darker areas indicate oil

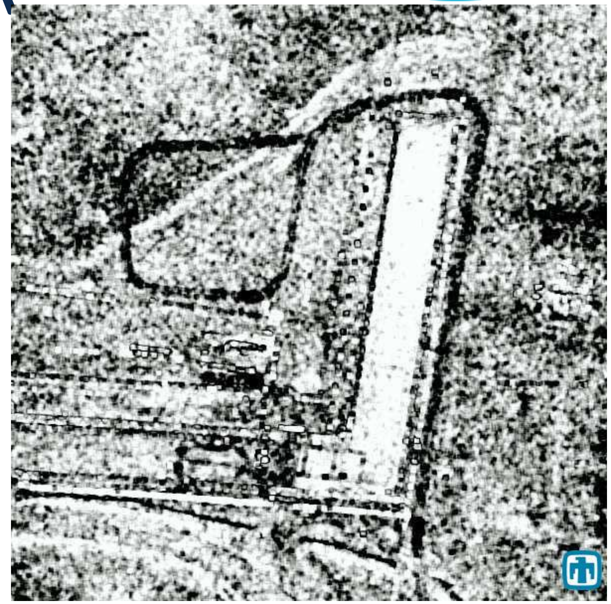


How SAR works

- 2-D Images made up of “range” and “azimuth”
- Range is line of sight distance
  - Determined by time between pulse transmission and receiving echo
  - Pulse width determines range resolution (smaller pulse, better resolution)
- Azimuth is perpendicular to range
  - Antenna length determines azimuth resolution
  - “Synthetic” aperture is distance plane flies to synthesize a long antenna
  - Azimuth position determined by doppler frequency of echos
- Lots of mathematics!



Change detection with SAR



- Take two images of the same scene at different times
- Register the images
- In areas where no change has occurred, the pixels in the two images are correlated, while in areas where change has occurred, they are uncorrelated
- Recent work
  - Creating products that help image analysts do their work quickly and efficiently

Sandia Directly Supports the Warfighter

Sandia, working with government partners, is providing technologies to the Warfighter which locate and help defeat improvised explosive devices (IEDs). These technologies are directly linked to saving the lives of many service men and women.

- Sandia's history of developing radars for Nuclear Weapons led to the development of advanced Synthetic Aperture Radars (SARs) that are today fielded on small unmanned aerial vehicles (UAVs) to locate and help defeat IEDs.
- This Counter IED system has been determined by the Department of Defense to significantly exceed all performance requirements and was unanimously recommended as a Proven Counter IED system.

- A Class III UAV outfitted with a Sandia developed SAR.

- Words from the Warfighter:

  "That thing is amazing, I wish we had it from the beginning a lot of people would still be around right now ….. I have witnessed 2 occasions in the passed few weeks where things could have turned out bad and you all saved the day."

Some other projects

- Reliability of photovoltaic systems
- Design of a ceramic separator for thermal batteries
- Quantifying uncertainty in measurements (position and velocity) determined from digital images
- Lots of data analysis
  - Margins and uncertainty
- Risk and reliability analysis
  - If we see parts that fail to meet the desired specifications, want to estimate the prevalence of the failures

Projects in Epidemiology
(Not at Sandia)

Sandia National Laboratories

- Modeling the occurrence of avian influenza in Indonesia
  - Highly pathogenic avian influenza (HPAI H5N1)
  - Endemic in domestic poultry
  - Can spread to humans and can cause death
  - Interested in when and where outbreaks are likely
- Investigating behavior of deer in Michigan, especially as it relates to the spread of bovine tuberculosis
  - Believe that wild deer spread disease to farm animals through shared feeding areas
  - Small number of deer responsible for large proportion of farm visits
- Joint work with Matt Farnsworth at the USDA

Internship Opportunities

Sandia National Laboratories

- Sandia has an extensive internship and co-op program, with summer and year-round options
- Several institutes
  - Enabling Predictive Simulation Research
  - National Security Engineering Institute
  - Physical Sciences Institute
  - Science of Extreme Environments Research Institute
- US Citizenship is required, 3.5 GPA, primarily engineering, computer or physical science, mathematics majors
- More information at www.sandia.gov/careers