Being a Statistician at Los Alamos National Laboratory

Emily Casleton (ecasleton@lanl.gov)

Deputy Group Leader, Statistical Sciences Group

Los Alamos National Laboratory

Approved for public release: LA-UR-22-20437

About Me

- Education:
  - BA in math/political science from Washington & Jefferson College
  - MS in statistics from West Virginia University
  - PhD in statistics from Iowa State University
- Attended a conference, met someone, had a summer internship at LANL, now they can't get rid of me.
- Current research: multi-source data fusion, Bayesian nonparametrics, Bayesian DOE for nuclear non-proliferation and Safeguards applications.
- Personal interests: moming, trail running, mountain biking, snowboarding, camping, hiking, backpacking.

Statistical Sciences Group, CCS-6

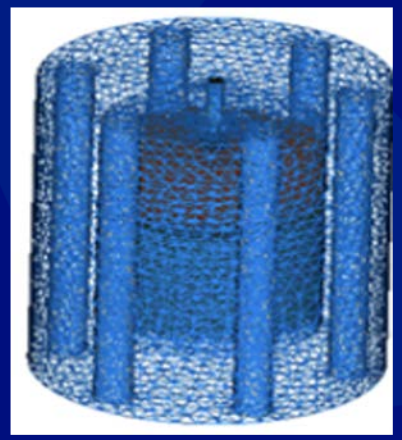
- 30 Statisticians
  - Almost all PhD in Statistics
  - 1 Post-doc
  - Between 5-15 students each summer
- Methods: We engage in research, development, and collaborative application of statistical methods.
- Applications: extend across a wide range of high-impact science, technology, and engineering areas supporting national and global security.

Direct Support of Weapons Science

- **Annual Assessment**: Certification to the President
- **Weapon simulation** through complex multi-physics codes; understand historic tests
- **Physical model understanding:** materials strength, thermodynamic equations of state, shock physics, turbulence
- **Nuclear forensics**: analysis of debris to answer questions
- **Pit Manufacturing**: Determine whether accuracy, precision, and overall uncertainty requirements are met

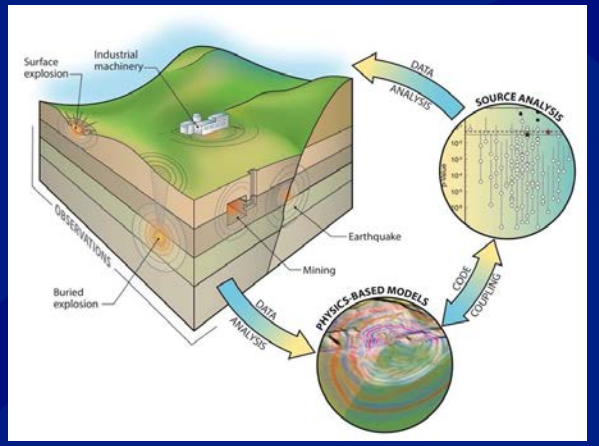
One Hop Away from Weapons Science

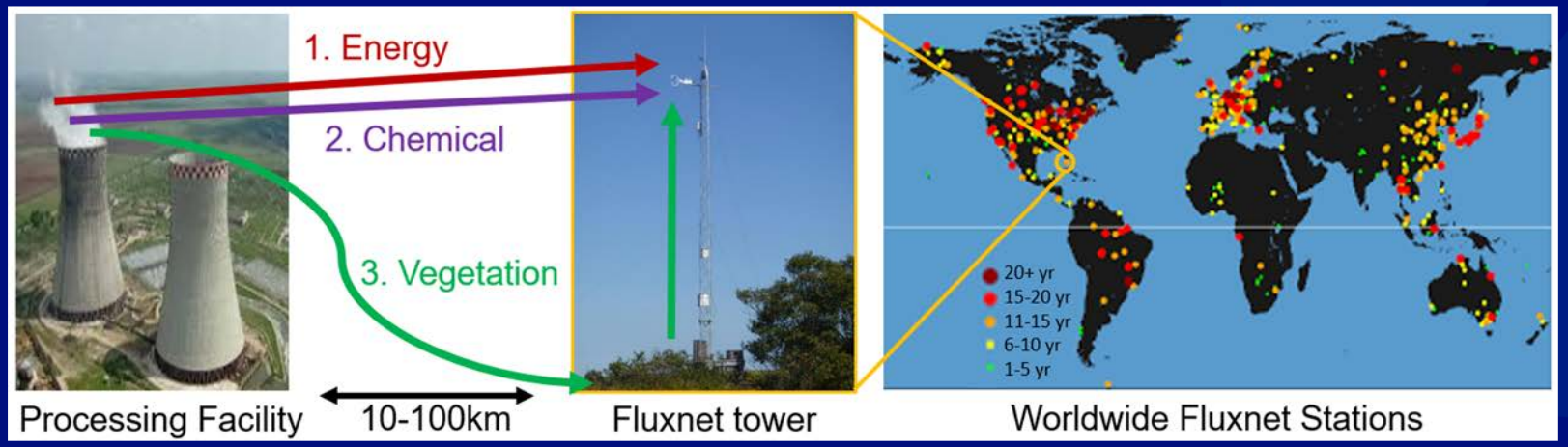
- **Nuclear Material Management**: Statisticians work with engineers and scientists to ensure the safe processing, transportation, storage and disposal of high-level nuclear materials
- **High Explosives development:** Understanding, measuring, and modeling explosive handling sensitivity
- **Nuclear Data:** identify sources of compensating errors in nuclear data and design validation experiments optimized to improve nuclear data estimates

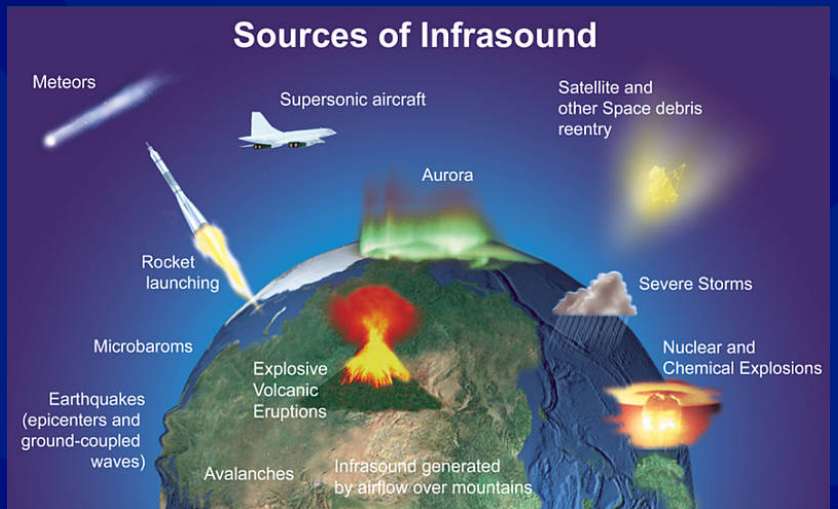


Nuclear Nonproliferation and Safeguards

- **Nuclear material control and accountability:** Nuclear material is a valuable and dynamic material. Use many resources to ensure nuclear material is safe.
- **Nuclear Non-proliferation:**
  - Using seismic or acoustic signatures
  - Using plants as biosensors
  - Through facility monitoring, patterns of life analysis

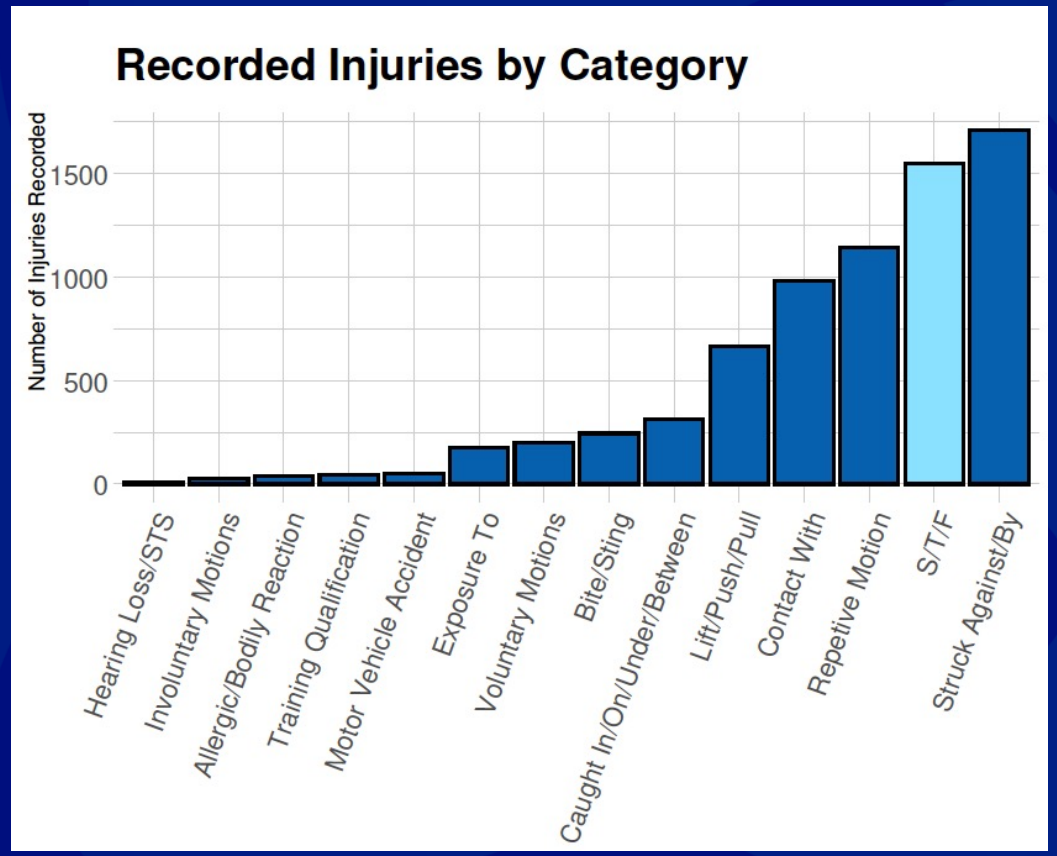






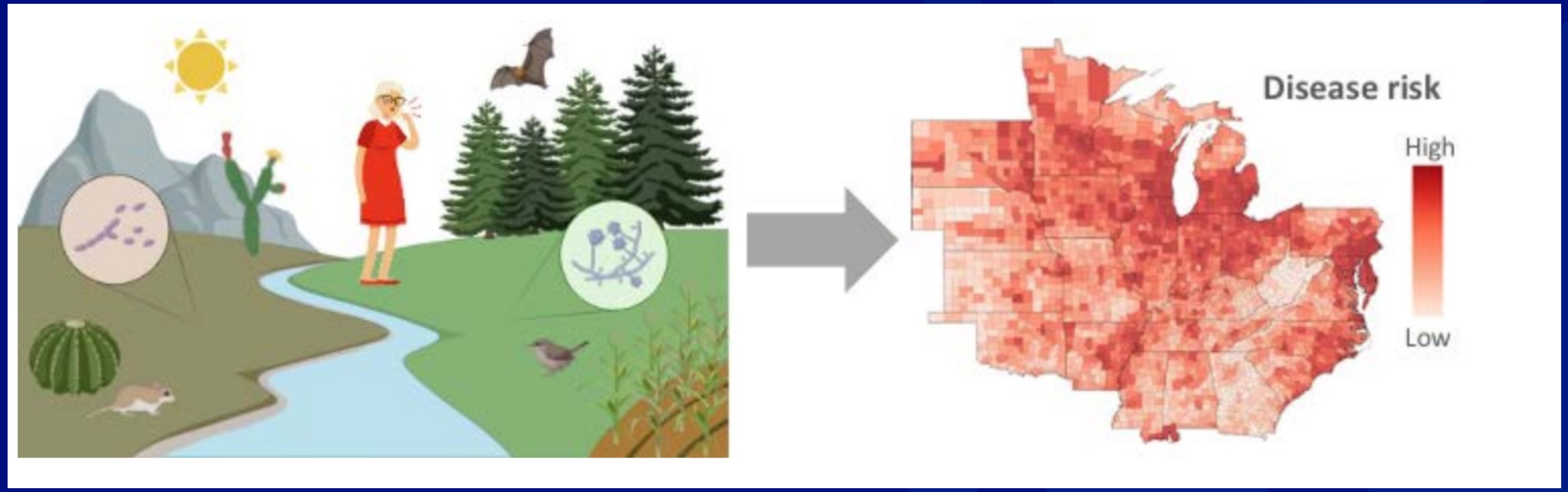
Analysis of the Analysts (and others)

- Use operational data streams to develop predictive capabilities using integrated data analytics for improved institutional decision-making
  - Colds and Flu, Covid-19
  - Vaccination Impact/Mitigation Strategies
  - Slips, Trips, and Falls
  - Telework Pilot Analysis
  - Where's Waldo (staff location info)
  - Risky Business (id key risk factors)
  - Student Career Tracker
  - Culture, Diversity, Performance Metrics
  - Facility Usage and Cost
  - Staff Retention Analysis



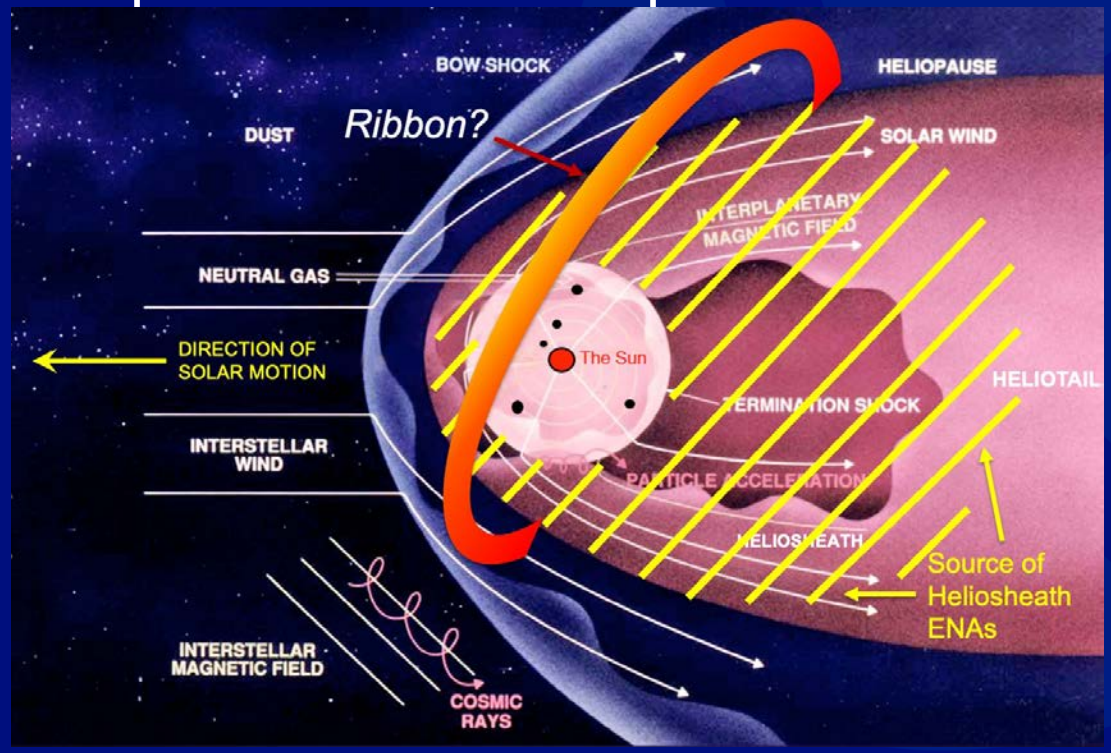
Really? LANL does that? Disease Modeling

Assess environmental and health impacts of airborne particulates-construct disease risk maps of histoplasmosis and valley fever



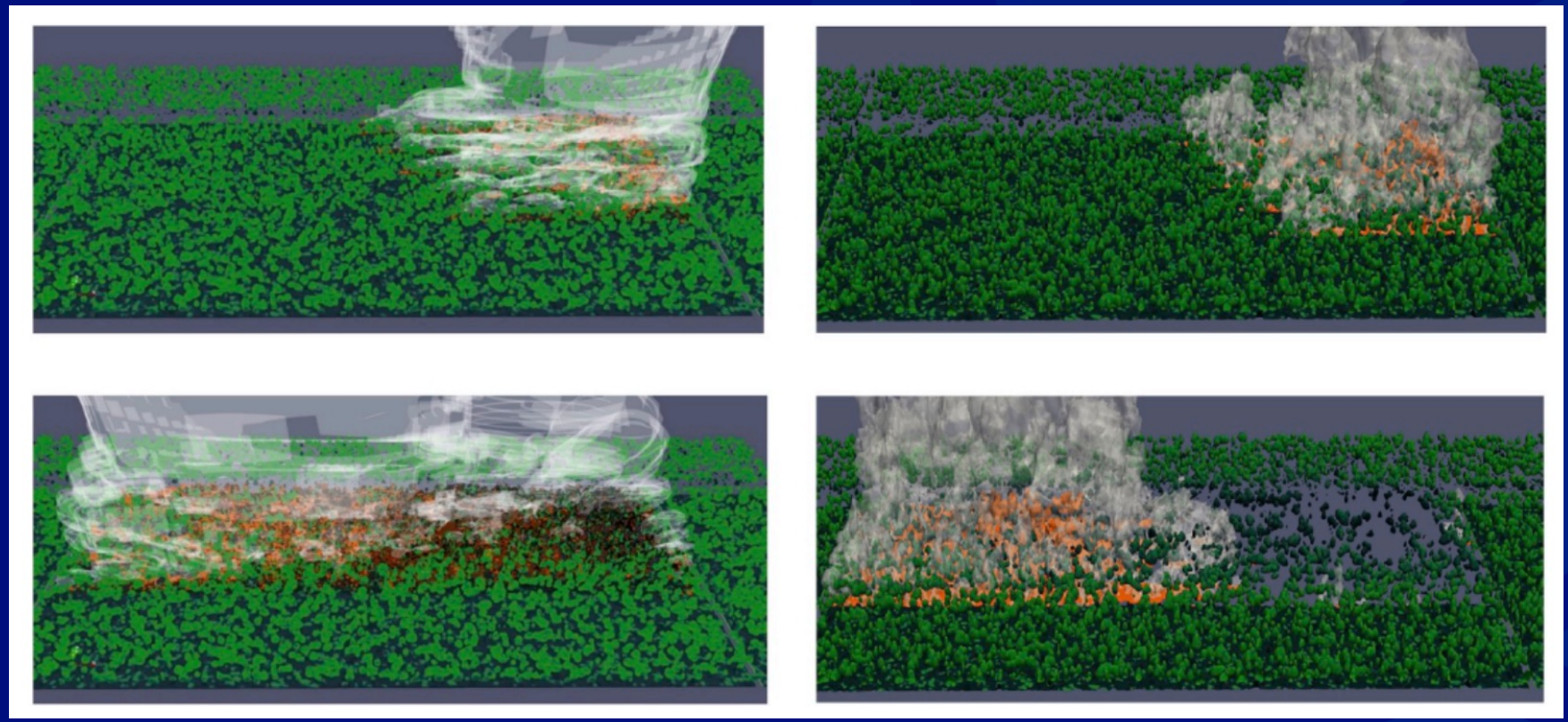
Really? LANL does that? Space Research

- NASA IBEX satellite mission has decades of data about the heliosheath – the boundary between our solar system and what lies beyond
- Data used to make maps of the heliosheath, which help discriminate between competing theories of aspects of the heliosphere.



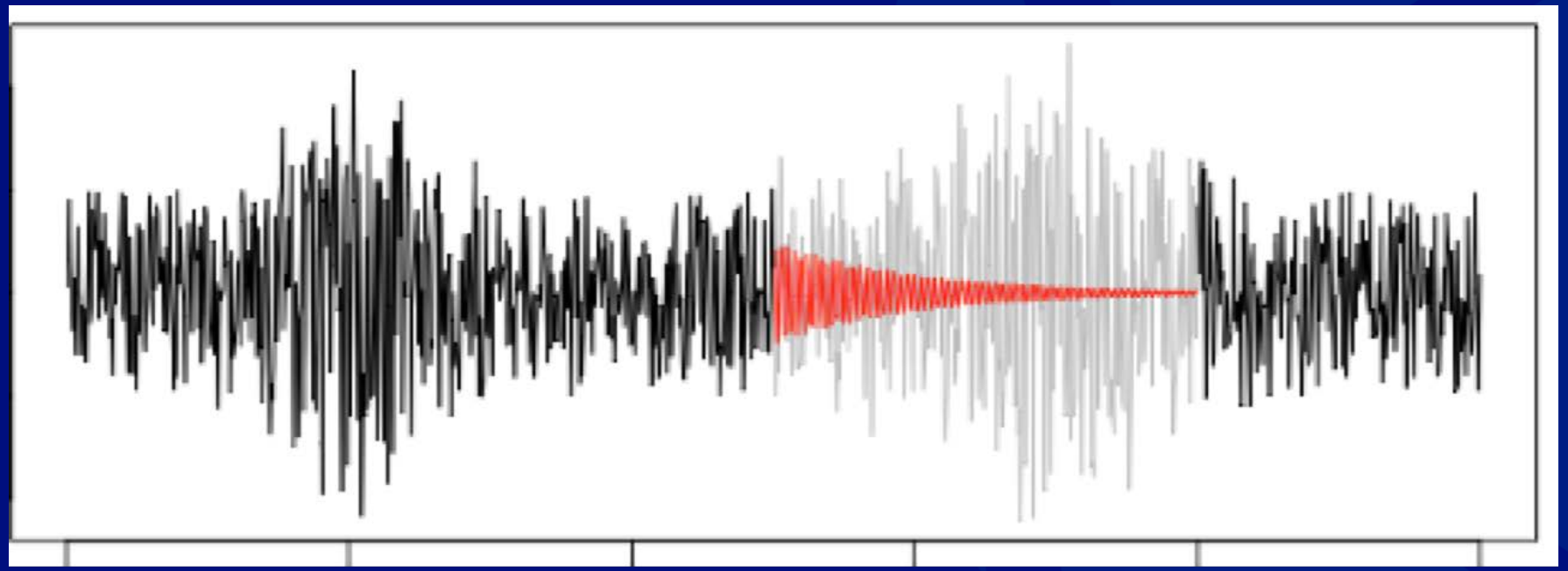
Really? LANL does that? Wildfire Simulations

Accurate heterogeneous structure of modeled forest fuels improves simulations of prescribed burns, for health, safety, and environmental and carbon management.



Really? LANL does that? Opioid Detection Device

Address the recent surge in opioid-related deaths in the United States, we are developing a nuclear quadrupole resonance device for safe, non-intrusive detection of synthetic opioids that leverages advanced statistical and machine learning techniques to recover target signals from noisy data



You can even do that?!?
Original statistical methodological development

Bayesian Multivariate Adaptive Regression Splines (BMARS)

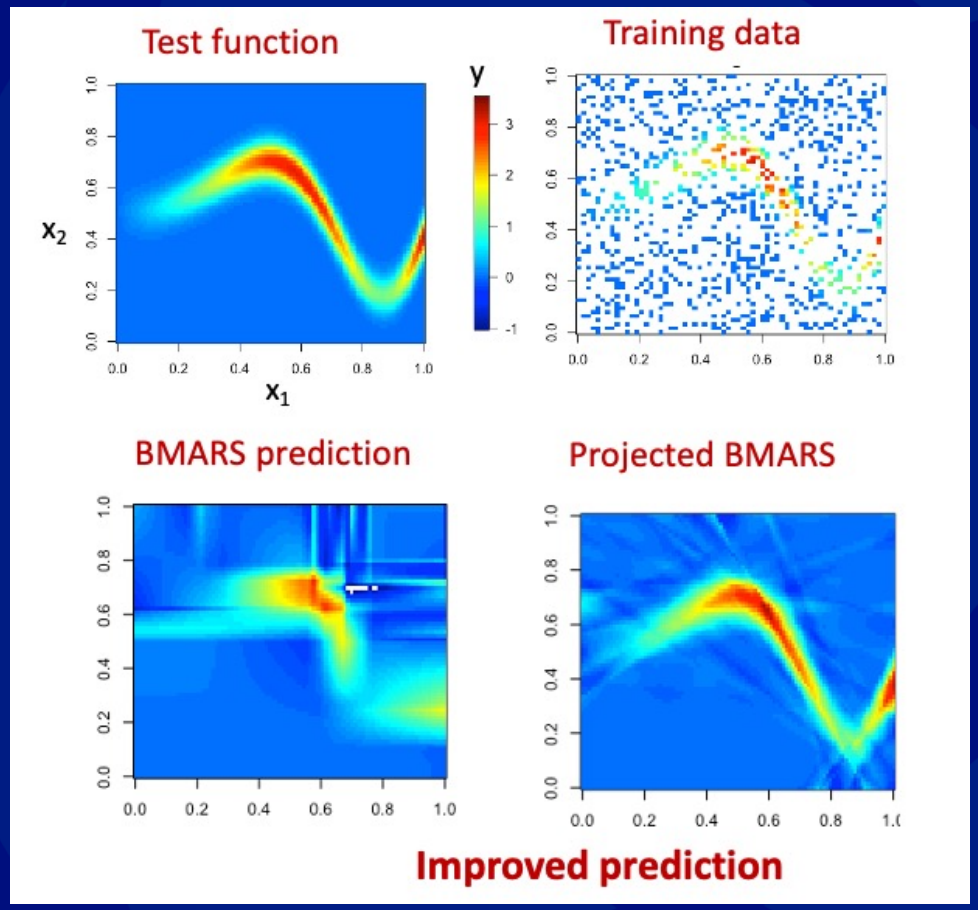
$$y = a_0 + \sum_{m=1}^{M} a_m \prod_{j=1}^{J_m} [s_{jm}(x_{v_{jm}} - t_{jm})]_+ + \epsilon$$

(J_m: degree of interaction; s_{jm}: sign; v_{jm}: variable; t_{jm}: knot)

***Projected* BMARS** model

$$y = a_0 + \sum_{m=1}^{M} a_m \prod_{j=1}^{J_m} [s_{jm}(\mathbf{x}'\mathbf{b}_{jm} - t_{jm})]_+ + \epsilon$$

(J_m: degree of interaction; s_{jm}: sign; $\mathbf{b}_{jm}$: projection; t_{jm}: knot)



Qualities we look for in an applicant

- Capable of research, development, and application of statistical methods
- Ability to collaborate with statisticians and subject-matter experts
- Good written and oral communication skills
- Strong statistical computing skills

Basically, statistical foundation with the ability to learn new things and be a good collaborator.

Other Details of CCS-6

- Active in external statistics community:
  - Journal reviewers and associate editors
  - Organizing conference sessions or whole conferences
  - Albuquerque chapter of the American Statistical Association
- Definition of “success” can be employee dependent

Los Alamos National Laboratory Is Looking for Excellent Statisticians

Work with Internationally Recognized Statisticians and Scientists

- ASA Fellows
- Journal Editorial Board Members
- Conference Founders
- A Network of International Collaborators
- 30 Colleagues, Early Career to Senior Professionals
- Over 40 Articles with 100+ Citations

Contribute to cutting-edge scientific applications of national importance in a multidisciplinary, collaborative environment.

Develop innovative methods to address the challenges of real-world data.

See your work having an impact!

For More Information Visit

stat.lanl.gov

Los Alamos NATIONAL LABORATORY EST. 1943

Backup Slides

i.e., more information

Statistical Methods in Support of Weapon Science

Broad Areas of Collaborative Problem Solving

Annual assessment

- Certification to the president: The stockpile is reliable.
- Tracking and trending to study aging effects

Weapon simulation

- Complex multi-physics codes
- Used to understand performance of historic tests
- Design capability for new systems

Supporting physical models

- Materials strength
- Thermodynamic equations of state
- Shock physics
- Turbulence, etc., etc.

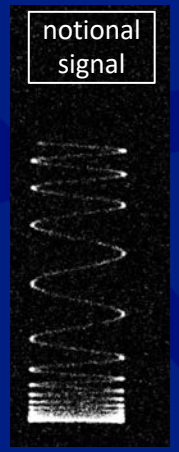
Nuclear Forensics

- Radiological analysis of debris
- Applied to historic U.S. tests: Did it go as designed?
- Preparedness for threats: What was it? Who did it?

Statistical Methods and Challenges

Data Challenges

- Nuclear testing ceased in 1992. First-hand experience is vanishing
- Diverse data types: photos, radiographs, oscilloscope traces, chemical assays, interference patterns, etc.
- Heterogeneous quality and systematics in historic data: e.g., sensor technology evolved



Calibration of physics codes

- Bayesian calibration – core capability
- Admit systematic discrepancies (code vs. observation)
- Hierarchical calibration across experiments from lab-scale to full-system

Extrapolation

- Often needed, never comfortable
- Confounding effects obscure aging trends
- Tough question: How conservative ought we be?

Data Analysis in Support of Pit Manufacturing Processes

Background/Problem

- Pit manufacturing processes require qualification of available analytical instruments and techniques

Objective

- Determine whether accuracy, precision, and overall uncertainty requirements are met
- Produce series of high visibility, high impact reports summarizing methods and findings

Current Actions

- Statistically analyze large amounts of data to establish if process qualification standards are met
- Employ a wide range of statistical methods to meet challenges and answer important questions

Takeaways

- Answering complex pit manufacturing questions with statistical methods is highly important
- This requires creative problem solving based on in-depth expertise of a wide range of statistical methods

Statisticians work with engineers and scientists to ensure the safe processing, transportation, storage and disposal of high-level nuclear materials

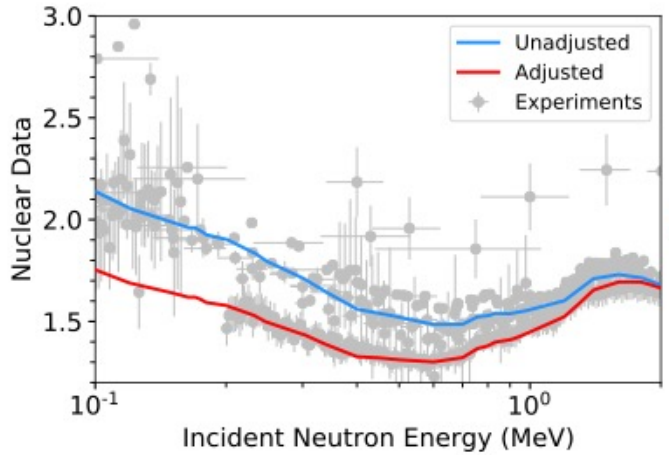
- Statistical approaches for problem solving : statistical-based surveillance sampling, image analysis, spatial analysis, experimental design, Bayesian analysis, stochastic predictive modeling of phenomena such as corrosion
- Other skills required : good written and oral communication skills, good collaborative skills – ability to communicate effectively with non-statisticians.
- Rich area for limitless challenging, important problems (*can even lead to a SPES award*).

2020 SPES citation - Department of Energy's (DOE) Waste Isolation Pilot Plant (WIPP) had a LANL drum rupture releasing radioactive material into the environment. Six months later, LANL experiments had not been able to recreate the energetic event causing the rupture. LANL experimentalists asked LANL statisticians to work on this mystery. Their collaborative work combined statistics, combinatorics, and nuclear chemistry to provide an explanation of the event. This collaborative work made an important contribution to the nation and to DOE's environmental management program that ensures the safety of the environment, public, and workers. Their work is reported in the article: E. J. Kelly, B. P. Weaver, D. K. Veirs, Quality quandaries: Statistical detective work to understand the isotopic ratios in drum 68660 and the radioactive release at WIPP, Quality Engineering, 29, 760-765 (2017)

ML-interpretability and Bayesian optimization to identify sources of compensating errors in nuclear data and design validation experiments optimized to improve nuclear data estimates

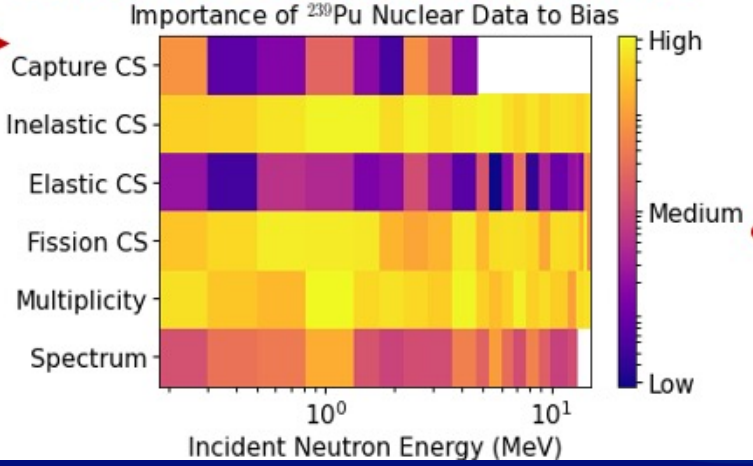
Neutron Transport Simulation (MCNP)

Validation Experiments

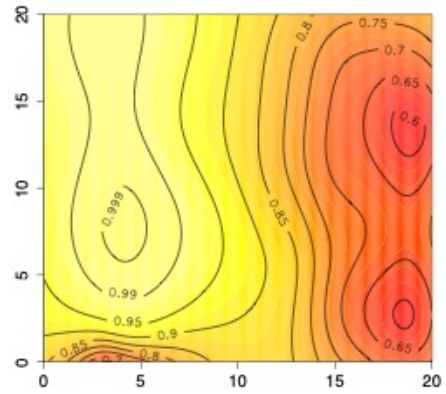


Experiment Refines Nuclear Data to Improve Simulations

ML-Augmented Search for Compensating Errors



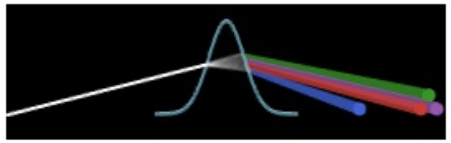
ML-Optimally Designed Experiment to Resolve Compensating Errors



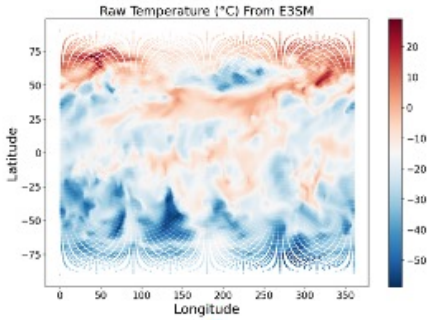
We perform distributed spatio-temporal inference *in-situ* while complex climate and space weather simulations are running

E3SM Energy Exascale Earth System Model

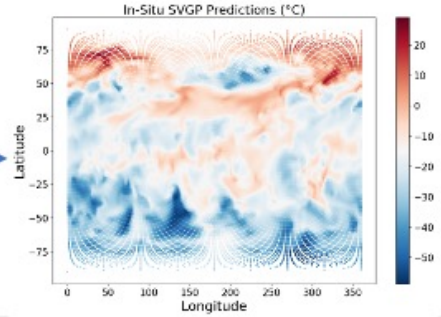
Launch Julia Instance



Evolve system



Communicate State to Julia for Inference



Estimate spatial regression of upper atmosphere temperature values

1 step of E3SM at this resolution takes 1.13 s. Fitting inference once per day (48 steps) means inference adds 7.5% to computation

Published: *Grosskopf, et al. ISAV (2021)*

Inducing Points Per Partition	MSE	Time-to-Fit (s)	Max Time-to-Fit (s)
2	0.968	3.688	3.926
5	0.257	3.846	4.057
10	0.181	4.283	4.566

Methodological Statistical Research: Novel Bayesian Nonlinear Regression Techniques

Bayesian Multivariate Adaptive Regression Splines (BMARS)

$$y = a_0 + \sum_{m=1}^{M} a_m \prod_{j=1}^{J_m} [s_{jm}(x_{v_{jm}} - t_{jm})]_+ + \epsilon$$

(J_m: degree of interaction; s_{jm}: sign; v_{jm}: variable; t_{jm}: knot)

***Projected* BMARS** model

$$y = a_0 + \sum_{m=1}^{M} a_m \prod_{j=1}^{J_m} [s_{jm}(\boldsymbol{x}'\boldsymbol{b}_{jm} - t_{jm})]_+ + \epsilon$$

(J_m: degree of interaction; s_{jm}: sign; $\boldsymbol{b}_{jm}$: projection; t_{jm}: knot)

- Projection allows us to put knots in non-axis-aligned space
- Inspiration: neural networks
- Projected BMARS can be viewed as a single hidden layer neural network, but we are learning the number of hidden nodes and the form of the (multivariate) activation function