NASA Perspective on Machine Learning

Nikunj C. Oza, Ph.D.
Leader, Data Sciences Group
nikunj.c.oza@nasa.gov

The Data Sciences Group at NASA Ames

Data Mining Research and Development (R&D) for application to NASA problems (Aeronautics, Earth Science, Space Exploration, Space Science)

Group Members

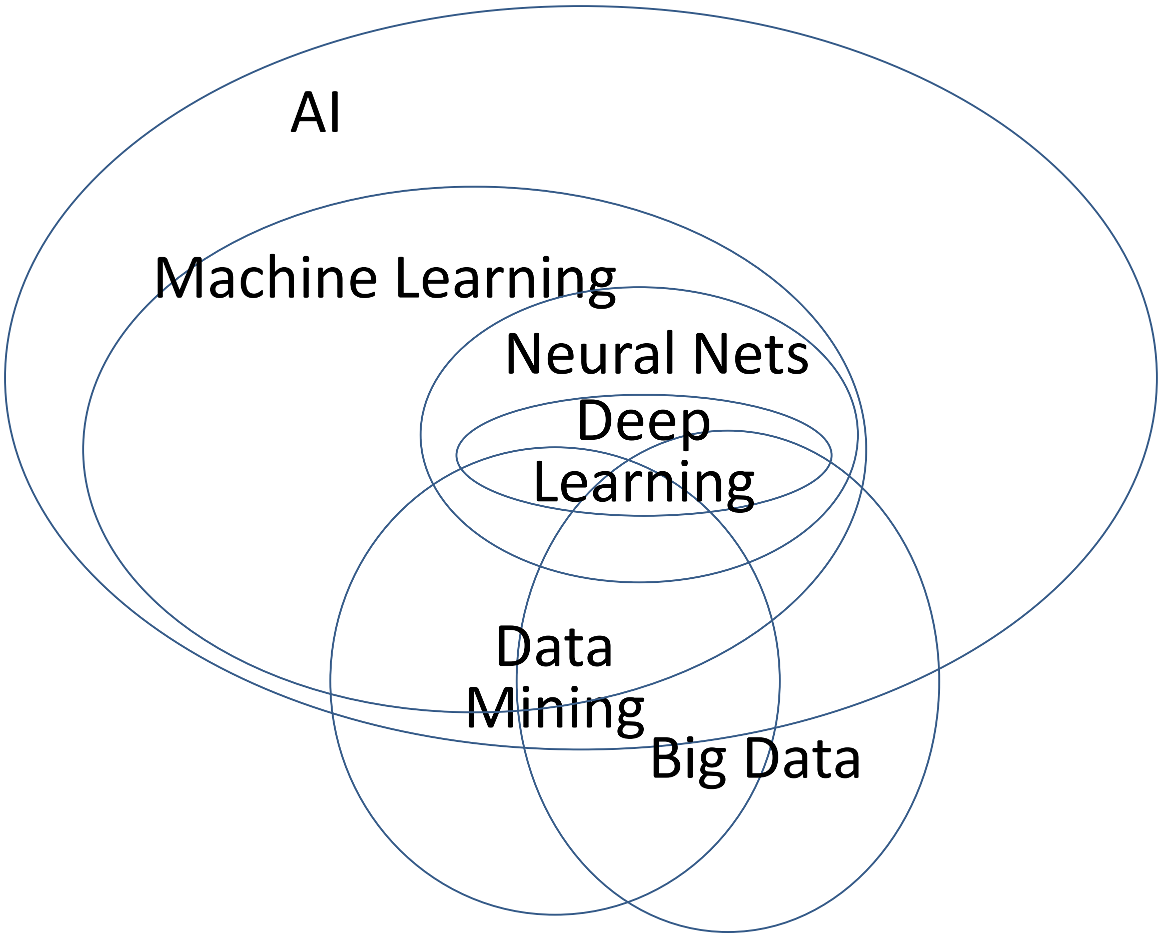
Ilya Avrekh
Kamalika Das, Ph.D.
Dave Iverson
Vijay Janakiraman, Ph.D.
Rodney Martin, Ph.D.
Bryan Matthews
Nikunj Oza, Ph.D.
Veronica Phillips
John Stutz
Hamed Valizadegan, Ph.D.
+ summer students

Funding Sources

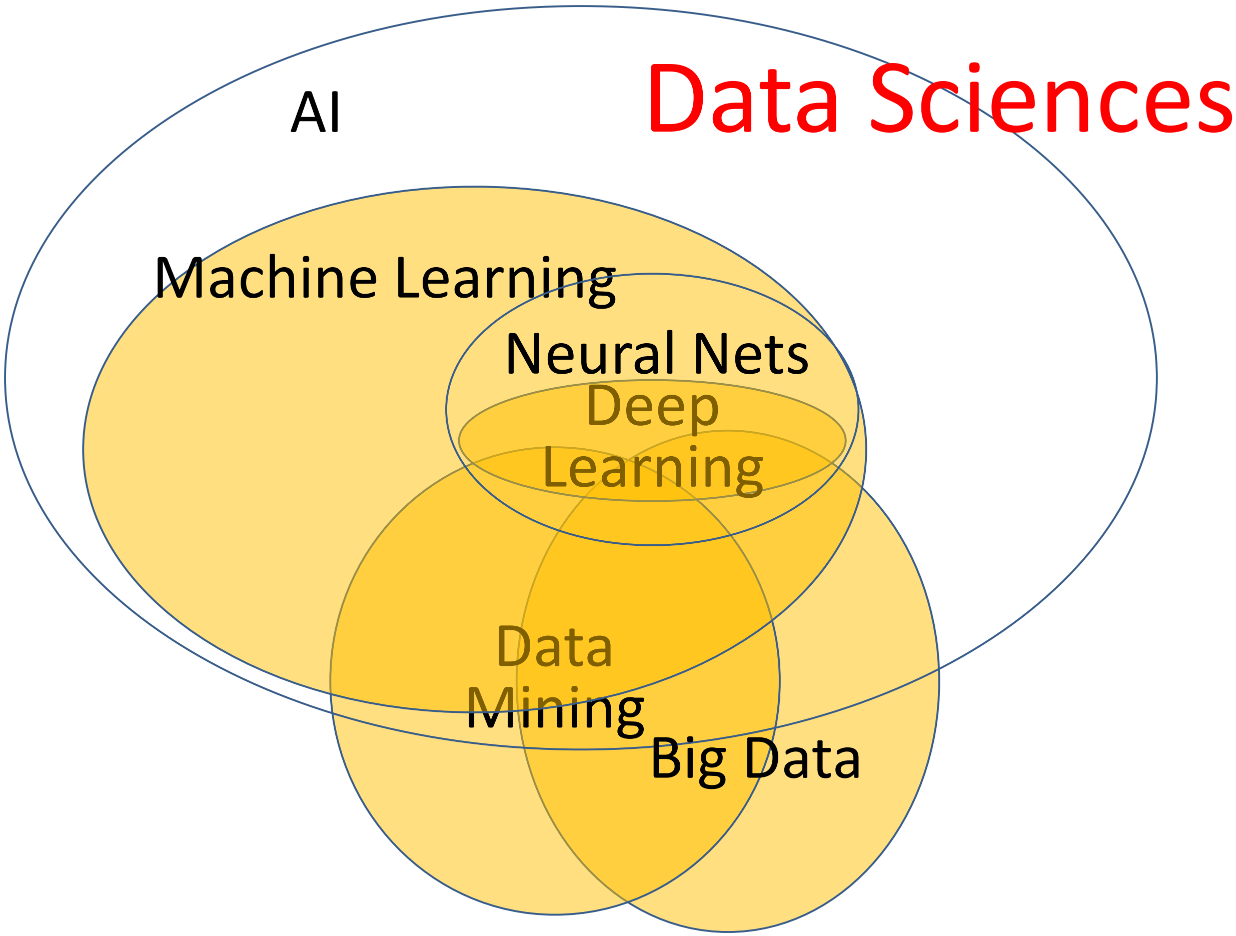
- NASA Science: AIST, CMAC programs
- NASA Aeronautics: ATD, SMART-NAS
- NASA Engineering and Safety Center
- NASA Human Space Exploration
- Aero seedling funds, Center Innovation Fund
- Non-NASA: DARPA, DoD

Team Members are NASA Employees, Contractors, and Students.

What is All This Stuff?



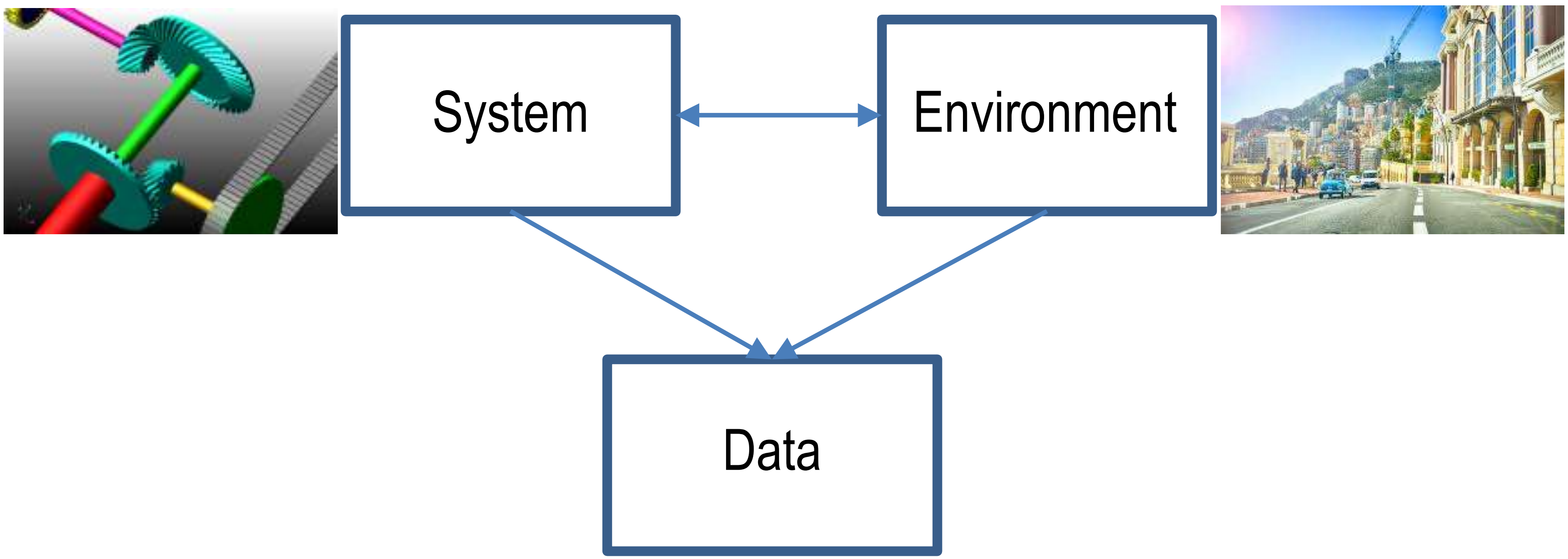
What is All This Stuff?



Collaborators

- Universities: Basic research in data sciences, domains
- Industry: Data sources, baseline methods, domain expertise
- NASA: Apply basic research, develop for NASA's needs, domain expertise, funding programs
- Other government: funding, domain expertise, data sources

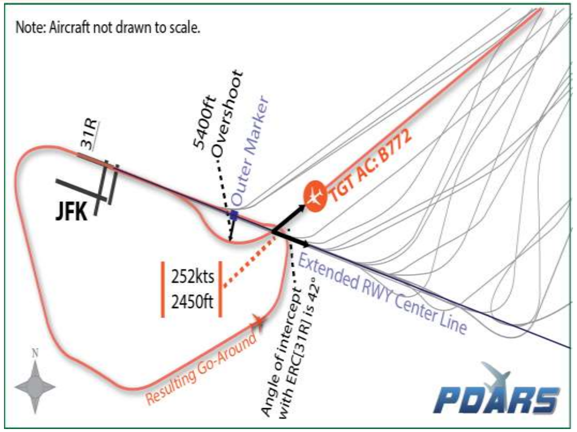
Data Sciences

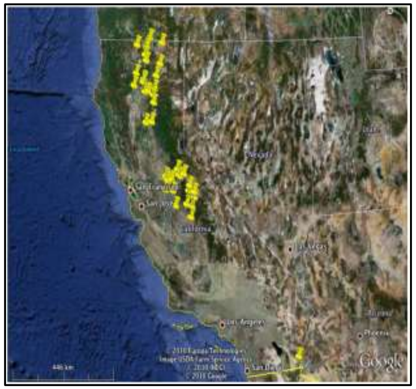


- Data are produced by system operating in an environment
- Data Sciences: Reverse-engineer system and environment
- Understand how system *really* works, correct system model errors, understand true impact of environment

Example NASA Machine Learning Problems

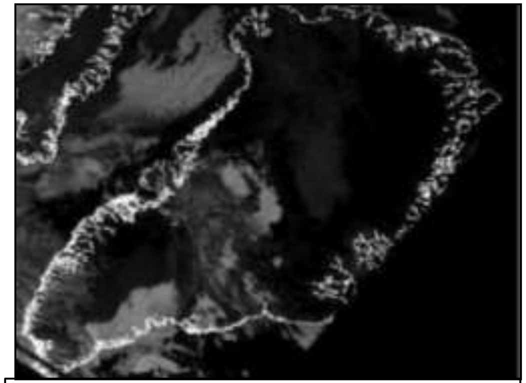
- Aeronautics
  - Anomaly Detection
  - Precursor Identification
  - text mining
- Earth Science
  - Filling in missing measurements
  - anomaly detection
  - teleconnections
  - climate understanding
- Space Science: Kepler planet candidates
- Space Exploration
  - system health management
  - astronaut health



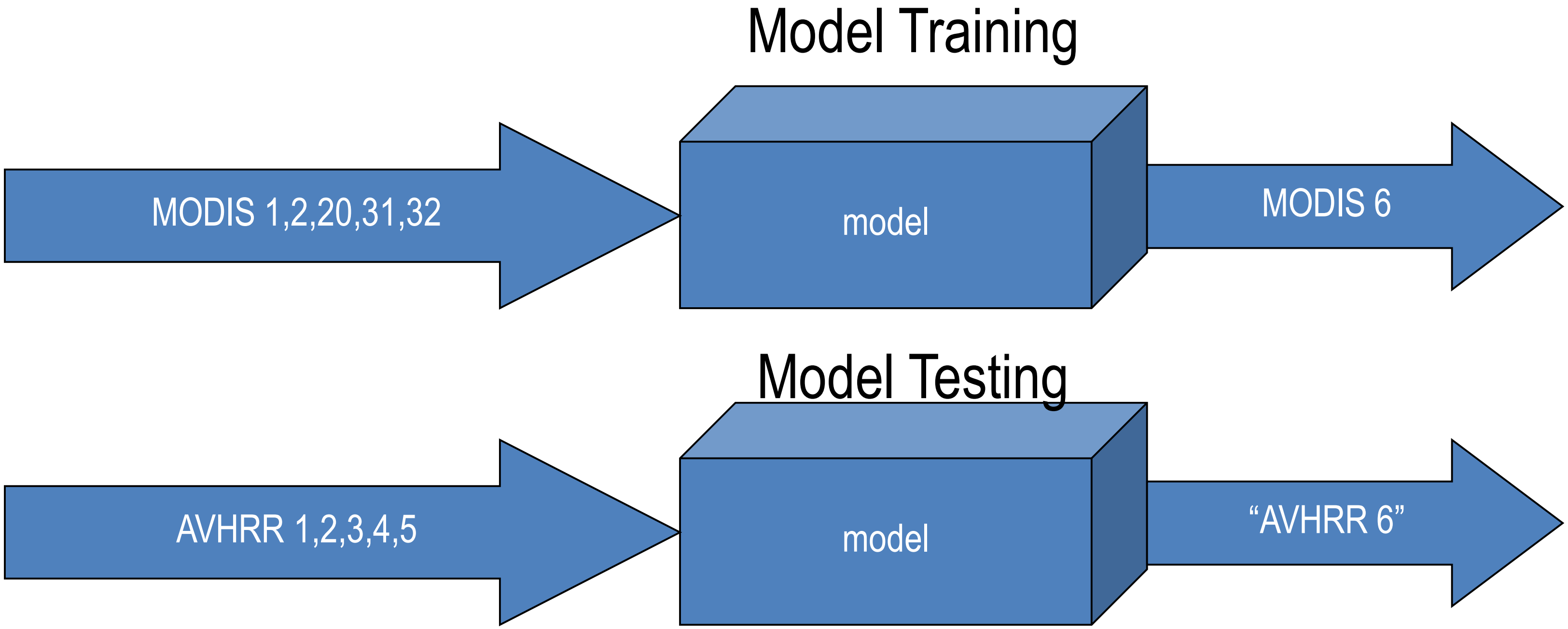


Data Mining for Earth Science Examples

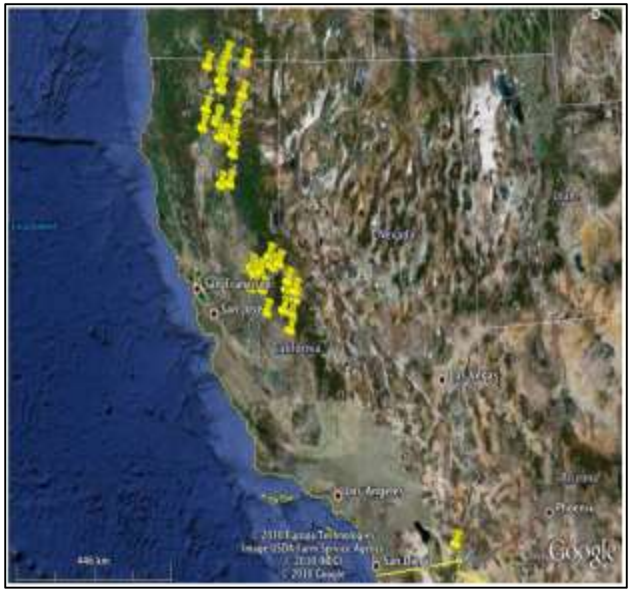
- Virtual Sensors
  - Regression to fill in missing or noisy sensor values, anomaly detection
  - Estimated MODIS channel 6 for older instrument (AVHRR)



Estimating MODIS channel 6 (useful for distinguishing clouds over snow and ice covered regions).



Data Mining for Earth Science Examples

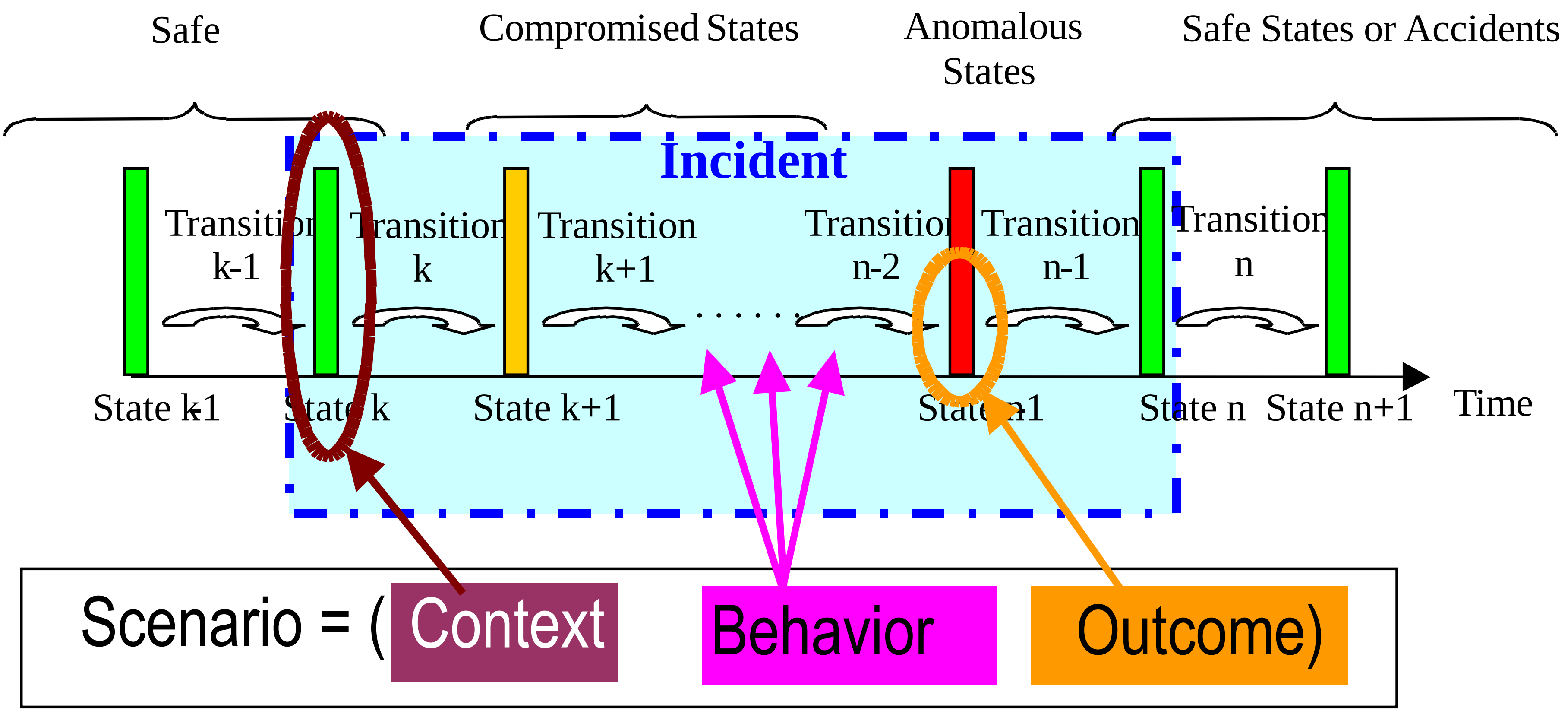


Top few outliers (yellow pins) identified by distributed 1-class SVM-based outlier detection algorithm in the California MODIS data.

Distributed Algorithms For Earth Science

- For large scale data where centralization is impractical
- Developed distributed 1-class SVM for anomaly detection.
    - 99% of the accuracy of centralized algorithm
    - 1% communication overhead running time (relative to embarrassingly parallel runs)

The Anatomy of an Aviation Safety Incident



From Irving Statler, Aviation Safety Monitoring and Modeling Project

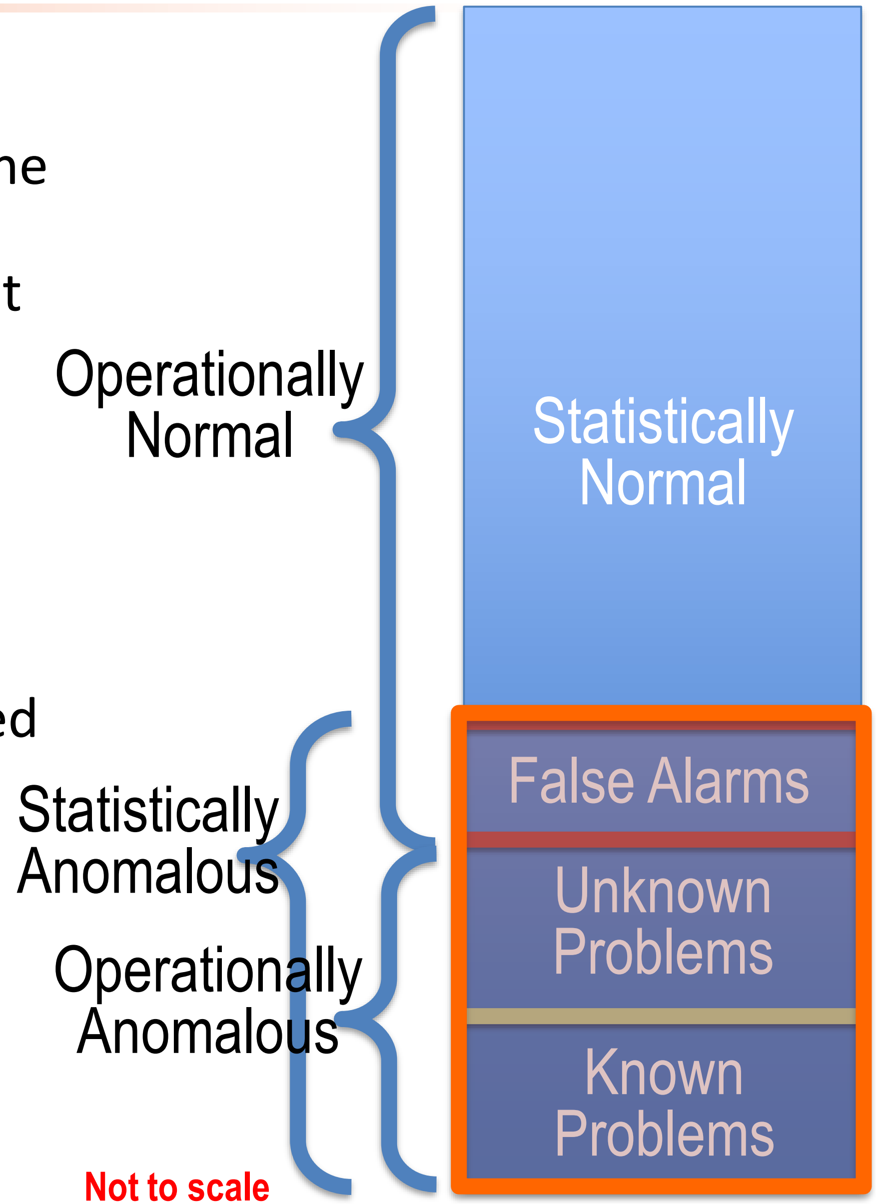
Current Methods of Finding Issues

Exceedance-Based Methods

- Known anomalies/safety issues
- Conditions over 2-3 variables (e.g., speed > 250 knots, altitude = 1000 ft, landing)
- Cannot identify unknown anomalies
- Low false positive rate, high false negative (missed detection) rate.

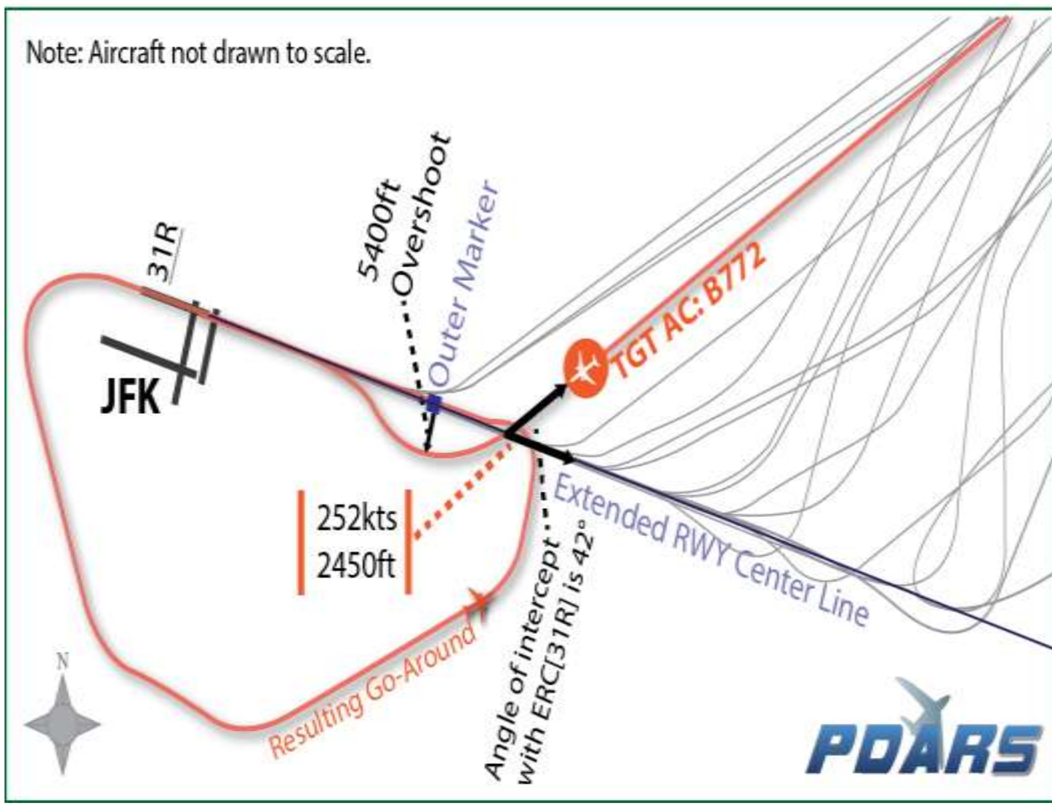
Data-Driven Methods

- DISCOVER anomalies by
  - learning **statistical properties** of the data
  - finding which data points do not fit (e.g., far away, low probability)
- Complementary to existing methods
  - Lower false negative (missed detection) rate
  - Higher false positive rate (identified points/flights unusual, but not always operationally significant)
- Data-driven methods -> insights -> modification of exceedance detection



High Speed Go-Around

- Overshoots Extended Runway Centerline (ERC) by over 1 SM
- Over 250 Kts @2500 Ft.
- Angle of intercept > 40°
- Overshoots 2nd approach



Four V's of Big Data

Volume[1]

- Radar Tracks: 47 facilities (1 year)
  - ~423 GB (Compressed)
  - ~3.2 TB (CSV)
- Weather and Forecast (Entire NAS)
  - CIWS ~2.8 TB

Velocity

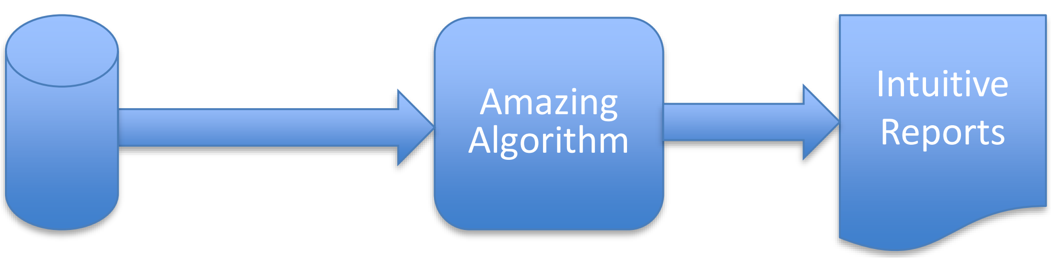
- Radar Tracks: 47 Facilities
  - ~35 GB/month (compressed).
  - ~268 GB/month (uncompressed)
- Weather and Forecast (Entire NAS)
  - CIWS ~233 GB/month

Veracity

- Data drop outs
- Duplicate tracks
- Track ending in mid air
- Reused flight identifiers

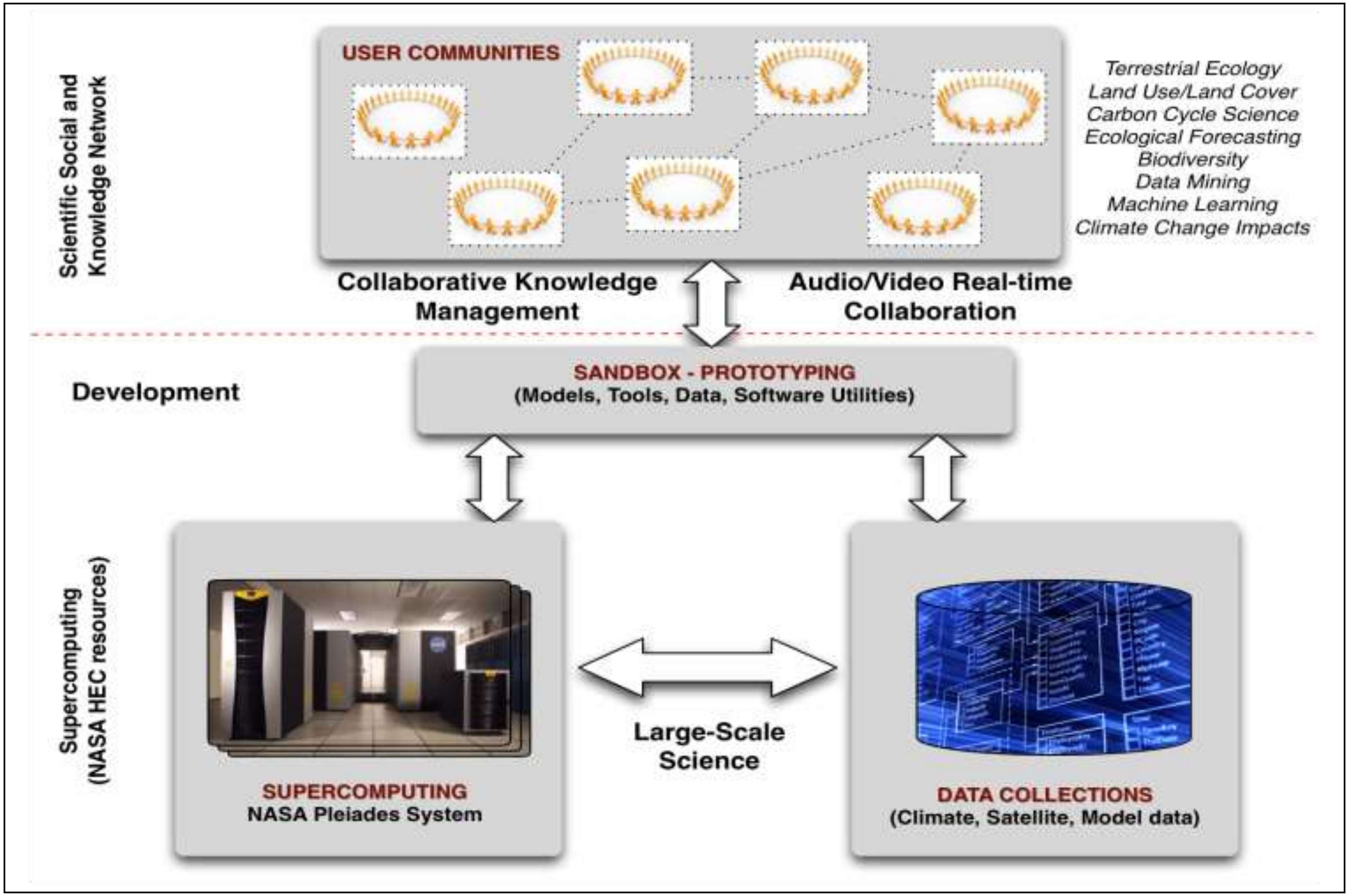
Variety

- Numerical (continuous/binary)
- Weather (forecast/actual)
- Radar/Airport meta data
- ATC Voice
- ASRS text reports (Pilot/Controller)

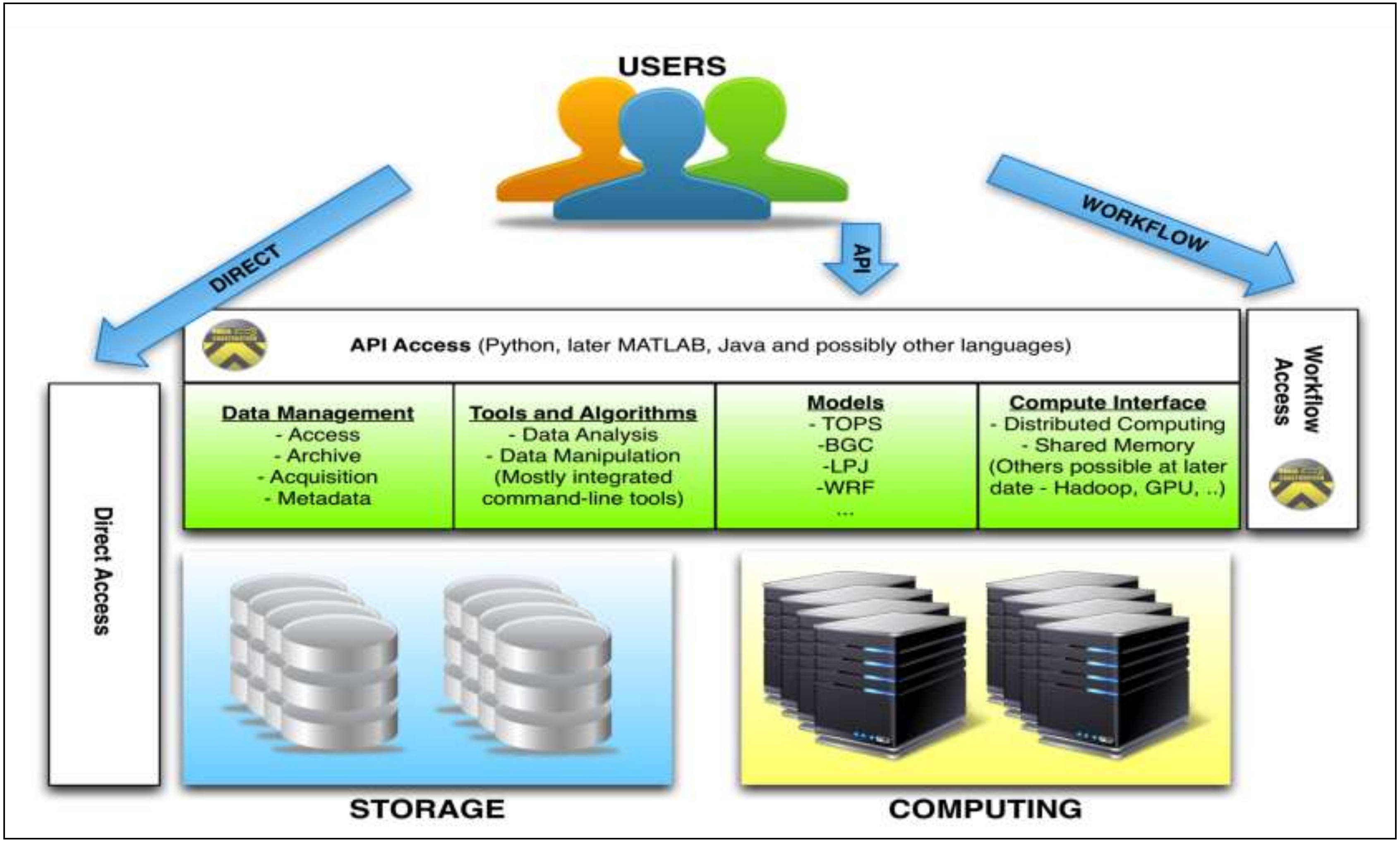


[1]But not always the right kind!

NASA Earth eXChange (NEX)



NEX Software View



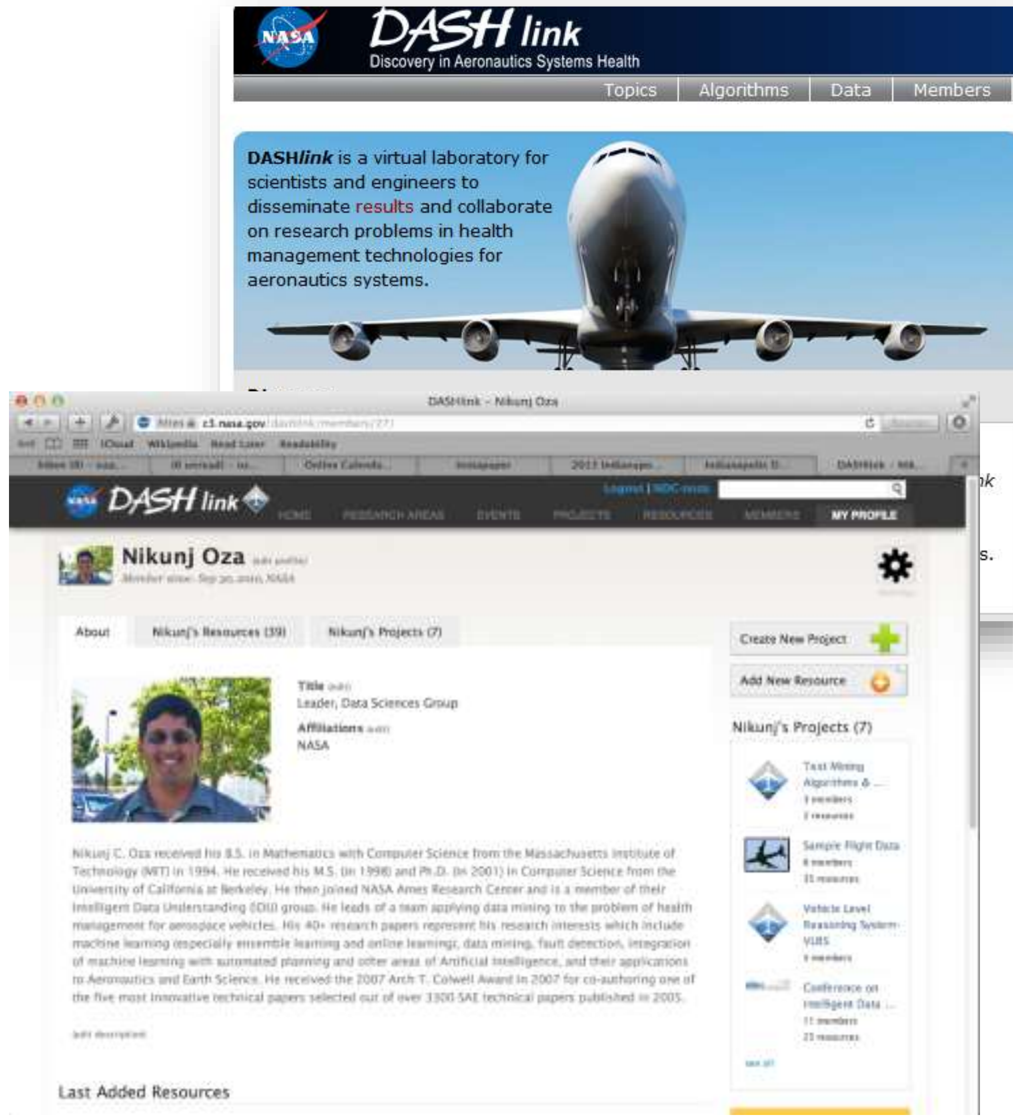
How do we get the Word Out?

DASHlink

disseminate. collaborate. innovate.
https://dashlink.ndc.nasa.gov/

DASHlink is a collaborative website designed to promote:

- Sustainability
- Reproducibility
- Dissemination
- Community building



Users can create profiles

- Share papers, upload and download open source algorithms
- Find NASA data sets.

Machine Learning Workshop - 2017

August 29-31, 2017

NASA Ames Research Center
Building 3
Moffett Field, CA

Peter Norvig

Vipin Kumar

SPEAKERS

- Peter Norvig, Keynote Speaker, Google
- Vipin Kumar, Keynote Speaker, Univ of Minnesota
- Bryan Matthews, NASA Ames
- Vijay Janakiraman, NASA Ames
- Deepak Kulkarni, NASA Ames
- Shawn Wolfe, NASA Ames
- Johann Schumann, NASA Ames
- Cliff Young, Google
- Guy Katz, Stanford University
- Damalika Das, NASA Ames
- Sangram Ganguly, NASA Ames
- Piyush Mehrotra, NASA Ames
- Kai Goebel, NASA Ames

TOPICS

- Machine Learning for Aeronautics
- Machine Learning for Human Space Exploration
- Program Synthesis for Efficient Machine Learning Algorithms
- Human Machine Interaction
- Machine Learning for Earth Science
- Machine Learning for Astrophysics and Planetary Science
- Supercomputing and Machine Learning

PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

Executive Office of the President
National Science and Technology Council
Committee on Technology

October 2016

THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

National Science and Technology Council

Networking and Information Technology Research and Development Subcommittee

October 2016

Ongoing and Future Work

- So far: desktop, HPC. offline, desktop
- Ongoing
  - in-time for online monitoring
  - Learning to improve analytics
- Future
  - Usability, portability for analytics deployments
  - embedded systems, autonomous systems
  - Use all platforms, in best way possible, on the fly

Thank You!

Contact: **nikunj.c.oza@nasa.gov**