Context
Data Science Life Cycle

Autonomy / Advisory Loop

- Quick Analysis
  - Quick Calibration
  - Quick QC
  - Quick Analysis
  - Quick Knowledge
- Annotate
- Advise / React
- Repeat

Quick-Look products enable earlier reaction
Advisory systems direct focus
What is Machine Learning?

Algorithms that inductively self-assemble from examples.

**Expert Examples**
(e.g. Estimates 1-10 of attractiveness of picture)

**Training Data**
(Encompasses domain of responsibility)

**Strength:**
Don’t specify rules
Don’t explain “how”

**Functional System**
- Interpretable as required
- Fixed code (V&V ready)
- Not ever-learning (mostly)
- EASILY UPDATED

Machine Learning simplifies & systematizes the building and updating of Autonomy / Advisory systems
Higher Level Questions / Actions

“Where should I start looking?”

“Show me more like this.”

“What is likely to happen next?”

“(Re)optimize my system.”

“How many kinds are there?”

“Show me the most interesting first.”

“What inputs are most informative?”

“Show me new things.”
JPL DS / ML Applications

1. Science-driven use cases
2. Explanability
3. “Let Me Help”
Holographic Life Detection

Drs. Lukas Mandrake, Gary Doran, Brian Bue

- Digital Holographic Microscopes
  - Big data (4D, ~GB/s), rare findings
- Motility ~ Life (composition agnostic!)
- HELM ML system detects, tracks, and classifies in messy, raw 2D holograms
AVIRIS-NG Hyperspectral

Dr. David Thompson, Dr. Brian Bue, et al

CH4 detection in four corners

Enabled ground team to find underground pipe leaks

Airborne Imaging Spectrometer

Multiple gas pipelines shut down / repaired

Machine Learning “That Matters”
Don’t Pre-Filter Data: No Data Quality Flags instead

Order by Trust Learn from Experts

New Approach: Ordering by Trust
No ground truth? No problem!

Physics
Signal Artifacts
Sanity Checks

What must be true to Trust Dataset?

Genetically discovers best, simple recipes to predict Trust.

Make Trust maps, Learn vars that predict Trust

Users take Trusted data first. Custom filtration & dataset size

Dr. Lukas Mandrake
Similar orbital tech used for
- Minerology maps
- Identifying crop-types
- Recognizing diseased citrus
- Estimating hurricane damage
Borup Fiord sulfur springs
Biosignature analog site for Europa
Detect and track from orbit (EO-1)
Support vector machine (SVM) classifier
26 detections as of May 31, 2016

EO-1 image (Hyperion)
Sulfur detection (yellow)

L. Mandrake et al., 2012. ACM TIST.
How to Supercharge NASA Data Science
The Growth of DS Concepts

- **Seedling Concepts**
  - $30-$50k
  - Proof of Principle
  - Single Problem

- **Demo System**
  - $150k-$300k
  - Extensive Validation
  - Single User Focus

- **Mission Adoption**
  - $400k-$1M
  - Mission Funding
  - Flight Readiness

- **Multi-Mission / Institutional**
  - $1M - $5M
  - Multiple Mission
  - Becomes Heritage

- **JPL has this under control**
  - R&TD system
  - Engineering Improvement
  - Data Science Working Group

- **Major bottleneck**
  - Can try for AMMOS Tech
  - Needed for approaching missions
  - Try to navigate science-based R&TD process

- **Missions ~receptive**
  - Require extensive validation for entry
  - Need to translate to “onboard” reqs

- **Each mission different “captain” of own ship**
  - Often afraid NASA won’t like Data Science / ML
  - Fear becomes main challenge

Shared Repositories for past DS / ML datasets & labels

Pot of Money to validate DS / ML systems to mission ready status

Mission AO specifically requesting new DS / ML techniques
Thank You!