Machine Intelligence towards Tomorrow’s Earth Science Data Systems

Manil Maskey

Program Officer - Earth Science Data Systems, NASA HQ

EARTH SCIENCE
DATA SYSTEMS

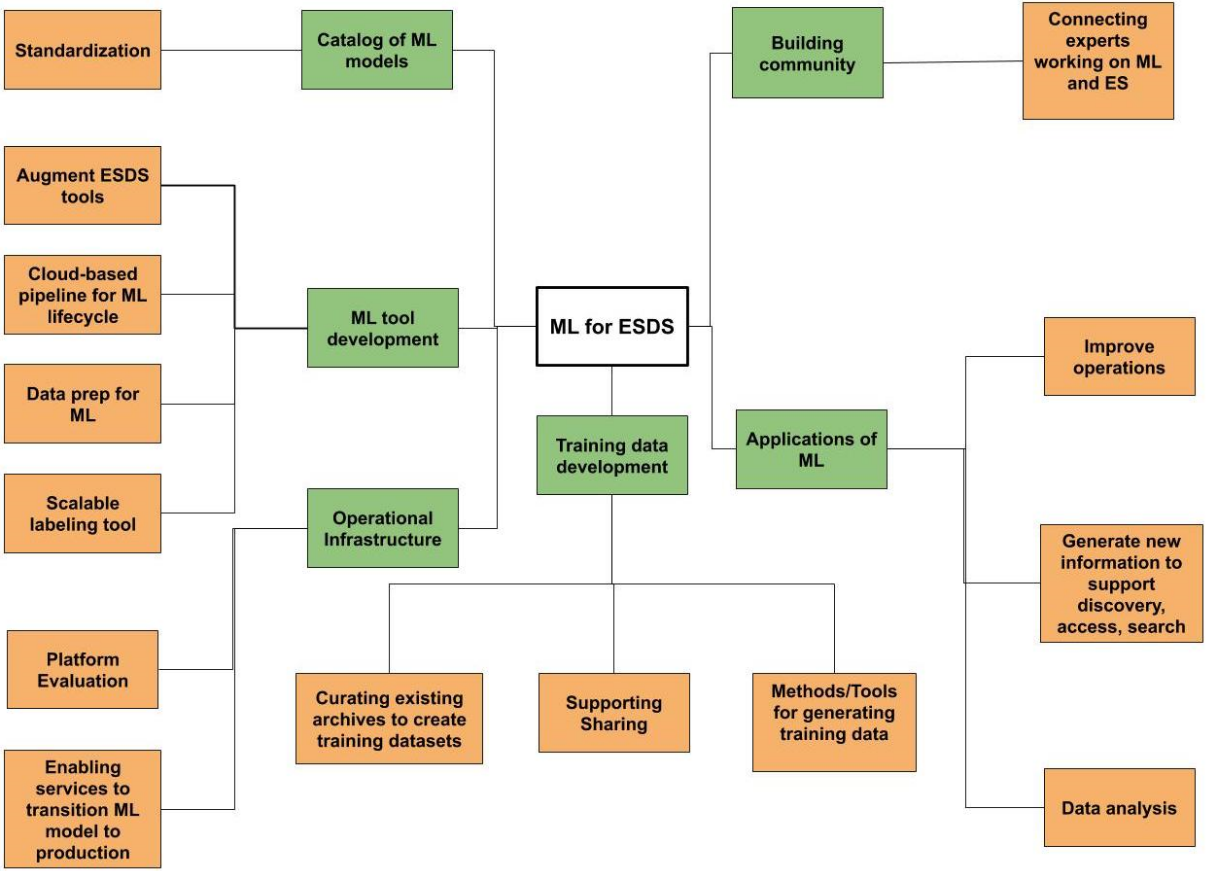
Why ML for Earth Science Data Systems

- Applying ML is critical to fully exploiting the large archives of Earth Science data
- Novel machine learning models show promise across disciplines by outperforming established benchmarks in prediction, forecasting, classification, and recommendations.
- Open and large datasets from multiple measurement systems now provide a unique opportunity to innovate new ways of analysis and obtain valuable insights.
- Mature ML systems can be applied to augment data systems operations, search/discovery, access, and use.
- Cloud platform now enables scaling ML applications with collocated data

Challenges in applying ML for Earth Science

- Issues with training dataset
  - Scarcity
  - Ensuring data access at a granular level
  - Inconsistency
  - Requires SMEs
  - Disparate systems/sensors
- Integration of heterogeneous data sources
- Use of suitable ML architectures and algorithms for a problem
- High barriers to entry (requires team science)
- Requires expertise in problem -specific optimization
- Lack of human -understandable solutions for interpreting ML model

ESDS Vision for ML



EARTH SCIENCE DATA SYSTEMS

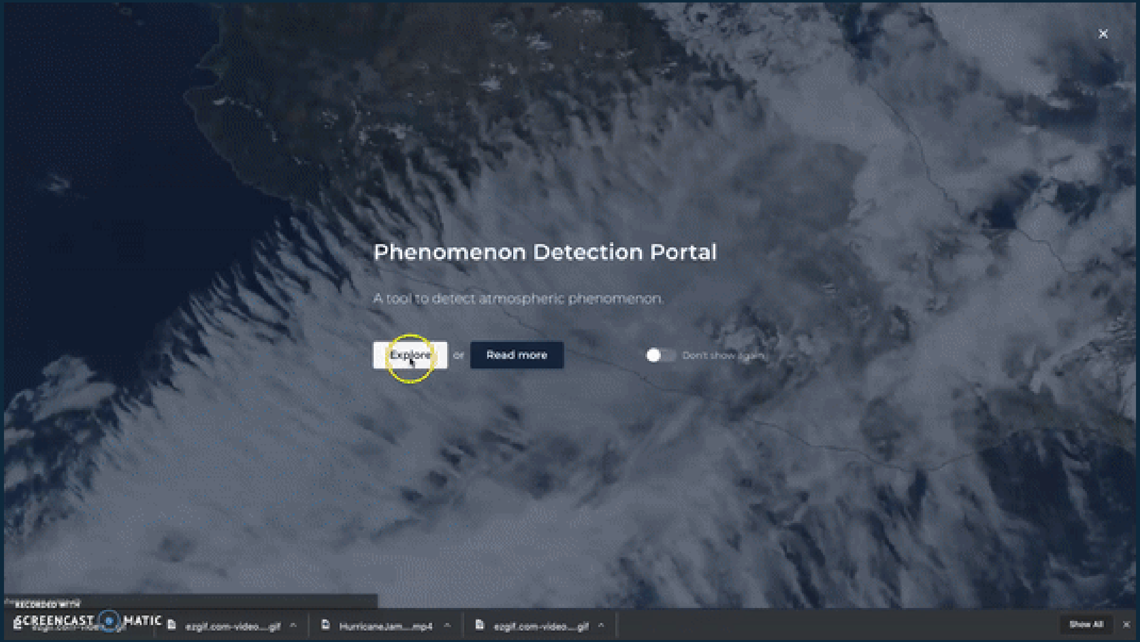
Near term priorities

- Training data generation/standardization/metadata
  - Workshop
  - Citizen science
- New Competed Solicitation
  - ML tool development
  - Training datasets
  - ML Applications
  - ML for selected mission datasets
- Building community
  - Workshop
  - Strategic Partnership

Ongoing efforts

- Applied science
- Improving operations
- Improving data discovery
- Transitioning research model to production
- Scalable labeling tool

Phenomena Detection Portal



Objective:

- Discover interesting events
- Use event space/time parameters to search for relevant datasets

Approach:

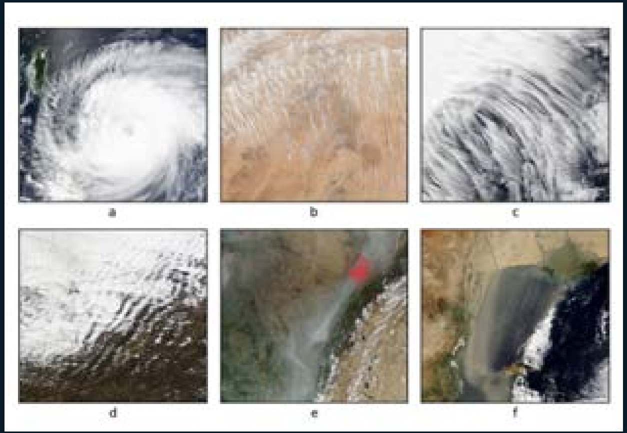
- Train DL algorithm with manually labeled dataset
- Run trained model daily on **GIBS** tiles
- Output identified events on portal
- Analyze events over space and time

Leverages Cumulus and GIBS

http://phenomena.surge.sh/

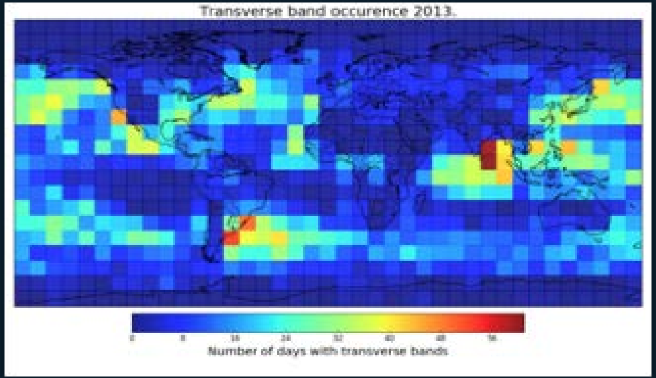
EARTH SCIENCE DATA SYSTEMS

Transverse Cirrus Bands Detection



Objective:

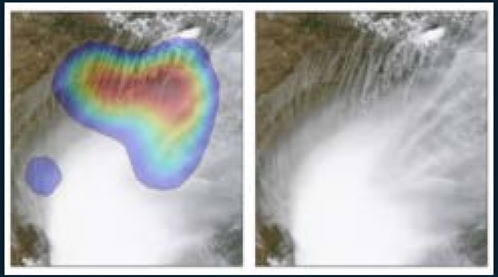
- Identify and study transverse cirrus bands which are associated with aviation turbulence yet very little is known about them



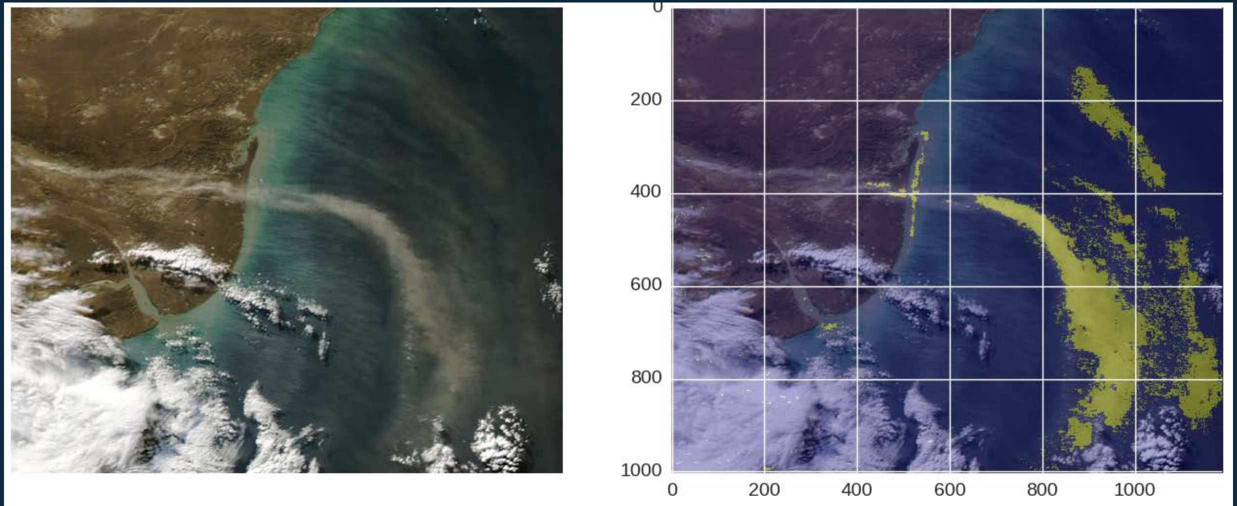
Approach:

- Use deep learning to identify bands in satellite imagery
- Analyze regions where these bands occur most often using GIBS imagery

~95% accuracy



High Latitude Dust Detection



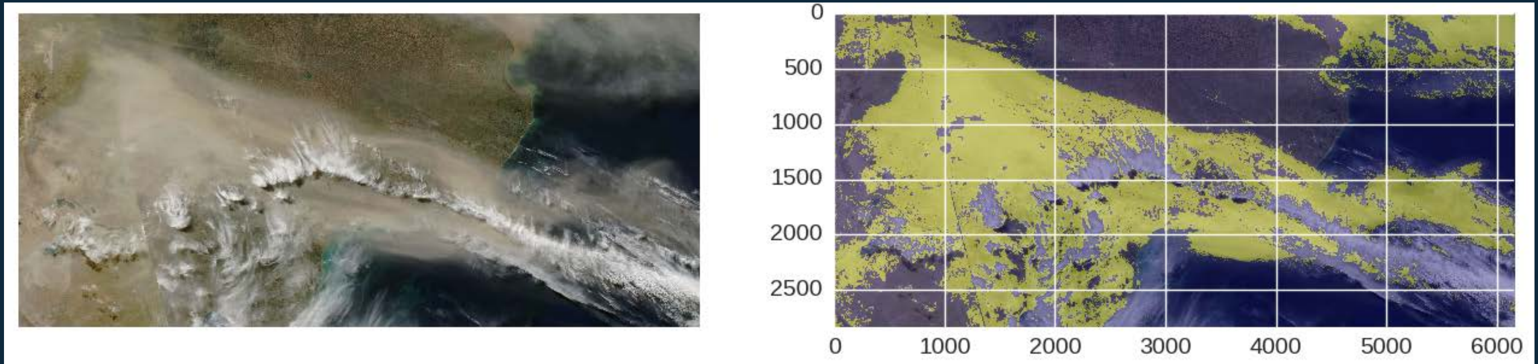
Objective:

- Develop a pixel-based High Latitude Dust detection

Approach:

- Train a DL-based pixel classifier
- Apply the classifier across GIBS
- Display results in phenomena portal

Leverages GIBS



EARTH SCIENCE
DATA SYSTEMS

Tropical Cyclone Intensity Estimation Portal

Objective:

- Use a Deep Learning (DL) method to objectively estimate tropical cyclone intensity
- Create a situational awareness tool to display estimation and other information in real-time

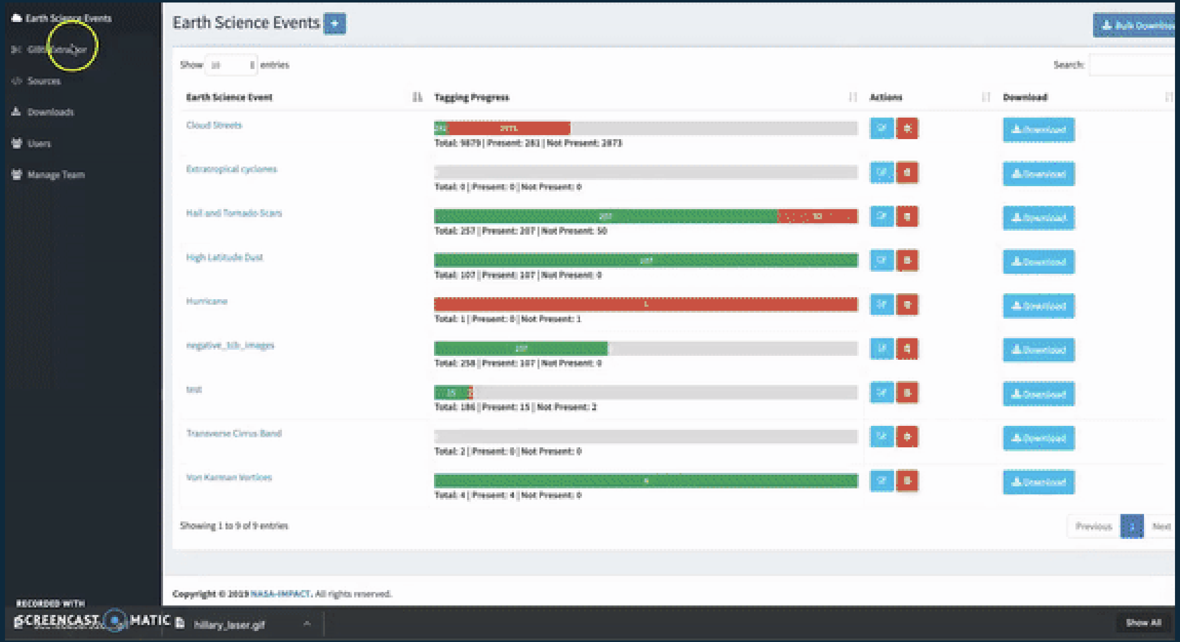
Approach:

- Train DL algorithm using heterogeneous datasets to estimate tropical cyclone intensity
- Use cloud infrastructure to build a real-time system triggered by event notification

Leverages Cumulus

http://hurricane.dsig.net/

Image Labeler



Objective:

- Develop a scalable image labeling tool for Earth science events

Approach:

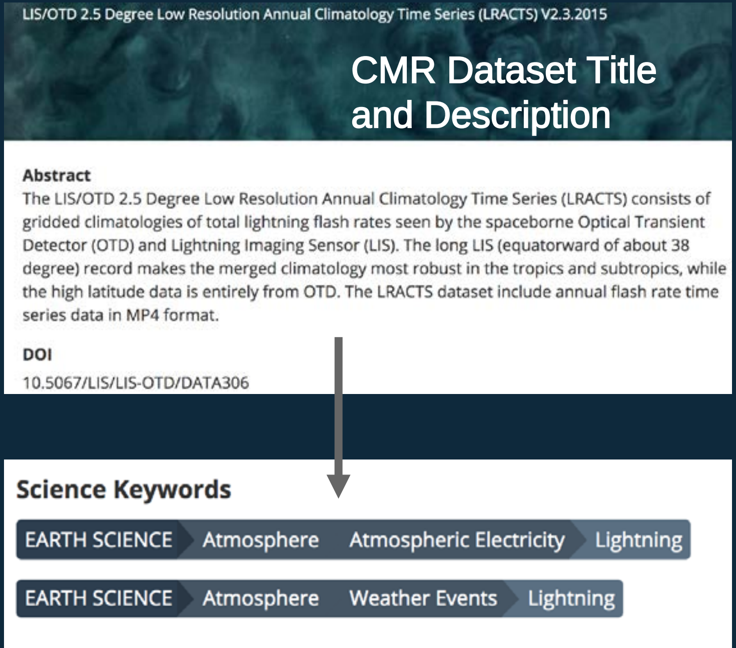
- Build a multi-user image labeler
- Integrate directly with GIBS
- Maintain a database of images for Earth science events
- Output data in ML ready structure

Leverages GIBS

https://labeler.nasa-impact.net

EARTH SCIENCE DATA SYSTEMS

Automated science keyword assignment



LIS/OTD 2.5 Degree Low Resolution Annual Climatology Time Series (LRACTS) V2.3.2015

CMR Dataset Title and Description

Abstract

The LIS/OTD 2.5 Degree Low Resolution Annual Climatology Time Series (LRACTS) consists of gridded climatologies of total lightning flash rates seen by the spaceborne Optical Transient Detector (OTD) and Lightning Imaging Sensor (LIS). The long LIS (equatorward of about 38 degree) record makes the merged climatology most robust in the tropics and subtropics, while the high latitude data is entirely from OTD. The LRACTS dataset include annual flash rate time series data in MP4 format.

DOI

10.5067/LIS/LIS-OTD/DATA306

Science Keywords

EARTH SCIENCE › Atmosphere › Atmospheric Electricity › Lightning

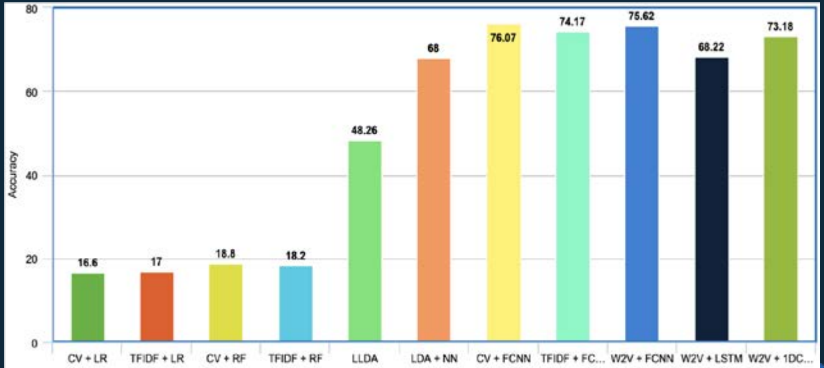
EARTH SCIENCE › Atmosphere › Weather Events › Lightning

Objective:

- Automatically tag metadata with science keyword.

Approach:

- Use CMR metadata to to train a classifier to select GCMD science keyword



~76% accuracy

EARTH SCIENCE DATA SYSTEMS

Building on existing portfolio

- Data/Metadata Archives
  - 27 PB of archived data
  - CMR: Rich curated content
- Cloud Infrastructure
  - Turn key ML stack
  - Pipeline for transitioning to production
- Tools
  - GIBS, Worldview, EONET,Cumulus
- Private industry partnership

Contact

Manil Maskey

manil.maskey@nasa.gov