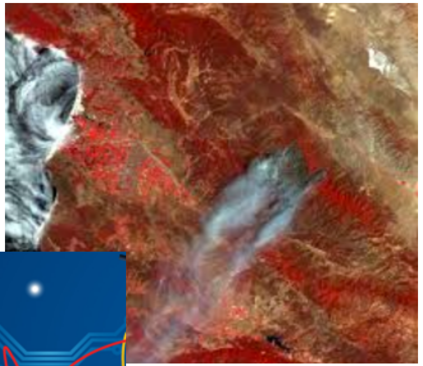
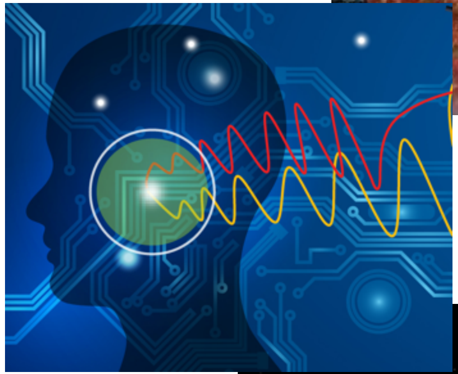
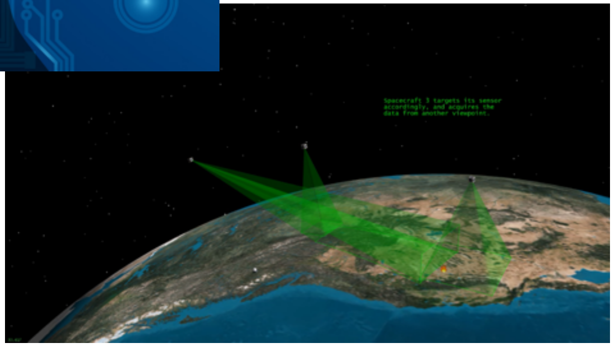
Artificial Intelligence for Advanced Earth Science Information Systems



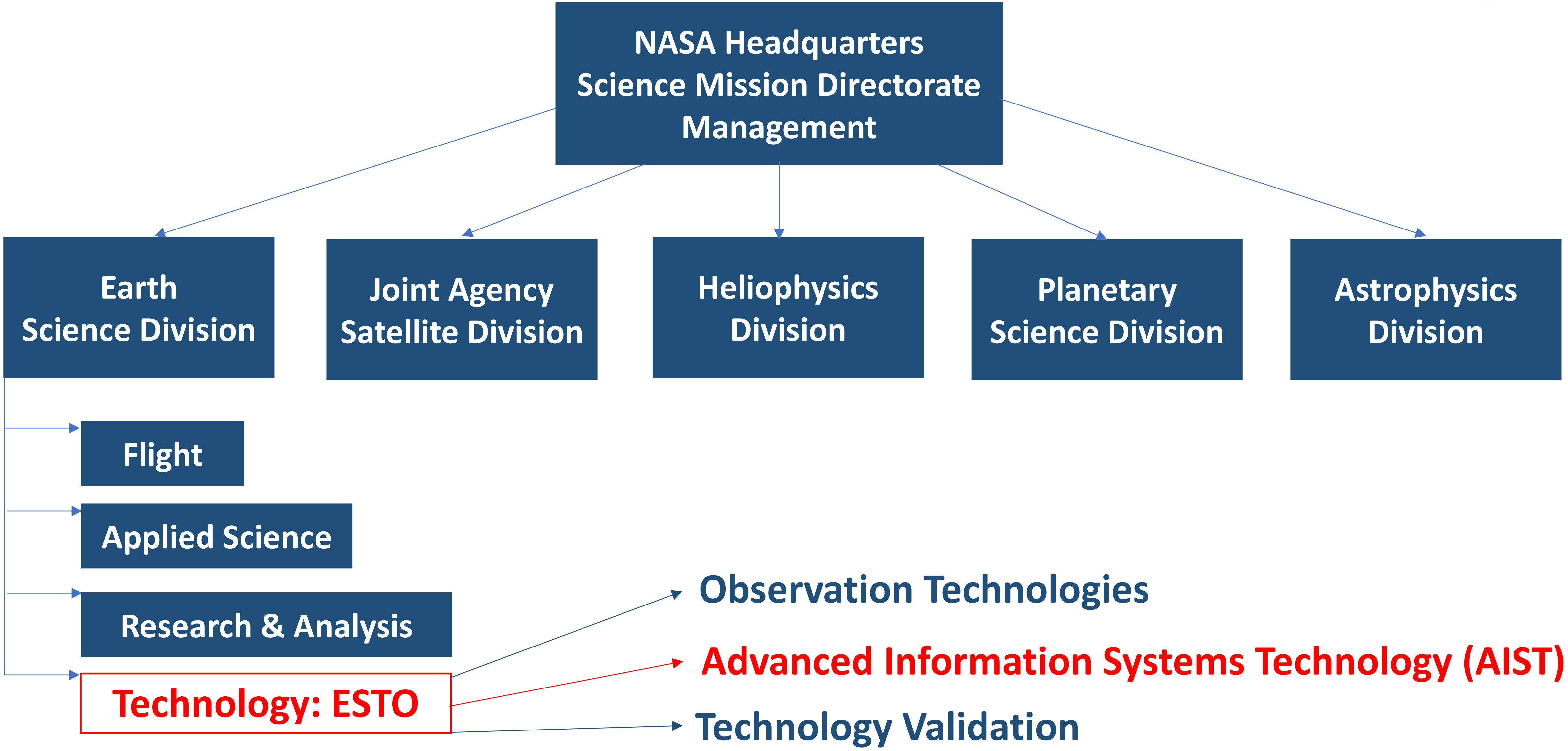




Jacqueline Le Moigne
Earth Science Technology Office (ESTO)
Advanced Information Systems Technology (AIST) Program

August 2020

Earth Science Technology Office (ESTO)

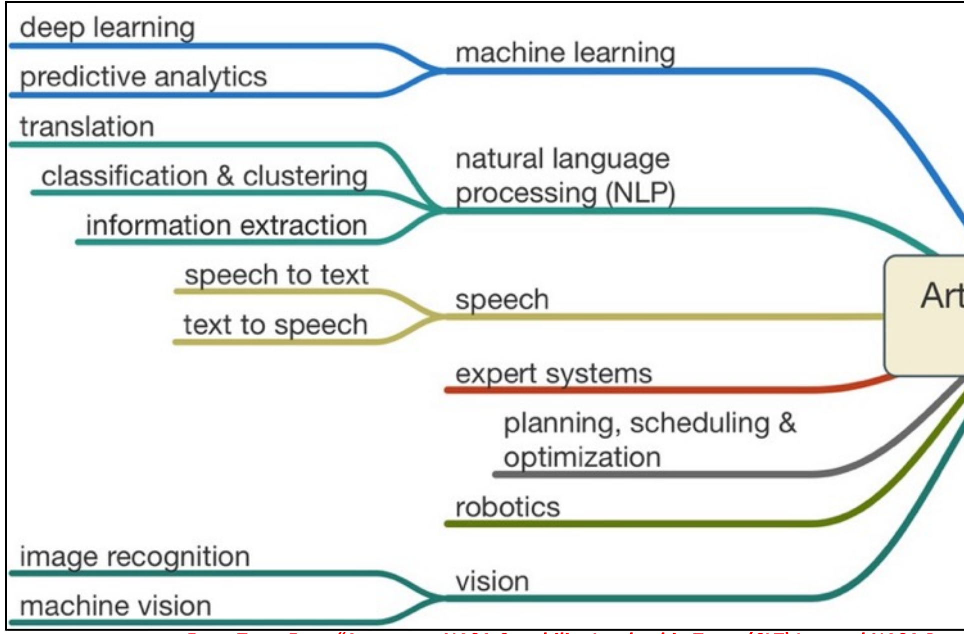


What is AI? Some Terminology

A few definitions

- *Artificial Intelligence (AI)* covers the development of the framework and of the technologies that enable a machine to perceive, reason, plan, act and learn both rationally and humanly.
- *Machine Learning (ML)* covers the sub-field of AI dealing with a machine capable of learning rationally and humanly.
- *Deep Learning (DL)* is a sub-field of ML dealing with very large Artificial Neural Networks including larger numbers of layers and of neurons, trained with massive amounts of data.
- *Machine or Computer Vision (CV)* is the ability to automatically understand any image or video based on visual elements and patterns.

- *Autonomy* is the ability of a system to achieve goals while operating independently of external control.
- *Automation* is the automatically-controlled operation of a system by mechanical or electronic devices that take the place of human labor – *Merriam-Webster*



*AI encompasses **many technologies** and **many applications***

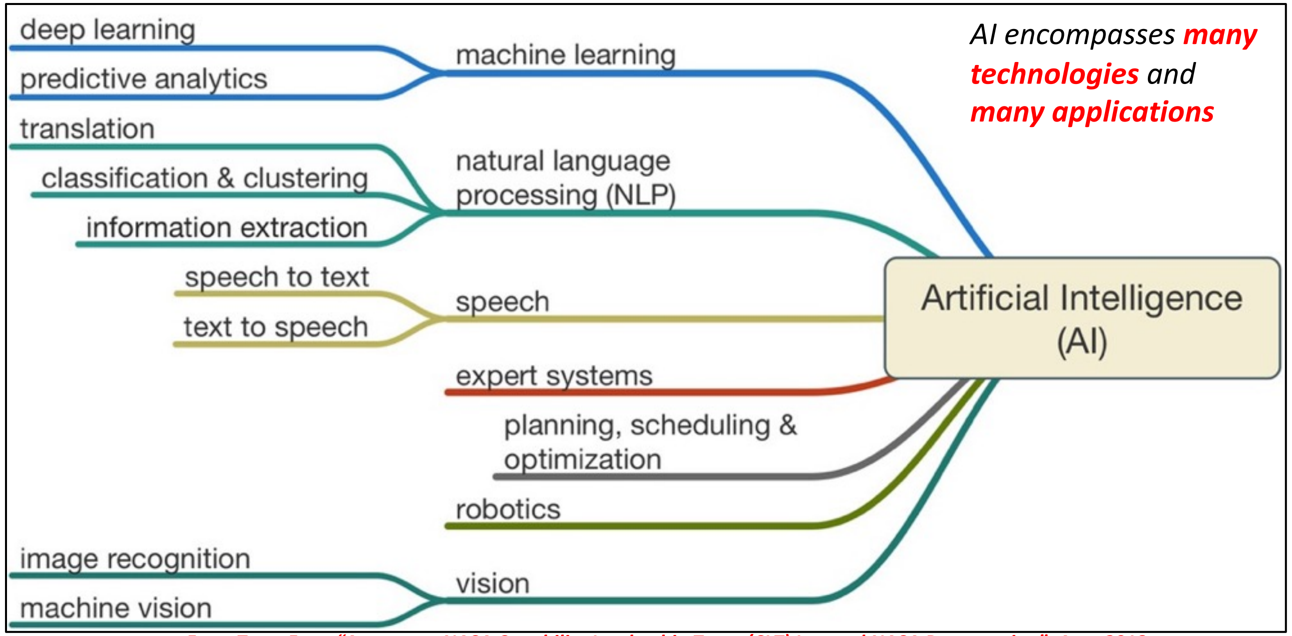
From Terry Fong "Autonomy NASA Capability Leadership Team (CLT) Internal NASA Presentation", Aug. 2018

What is AI? Some Terminology

A few definitions

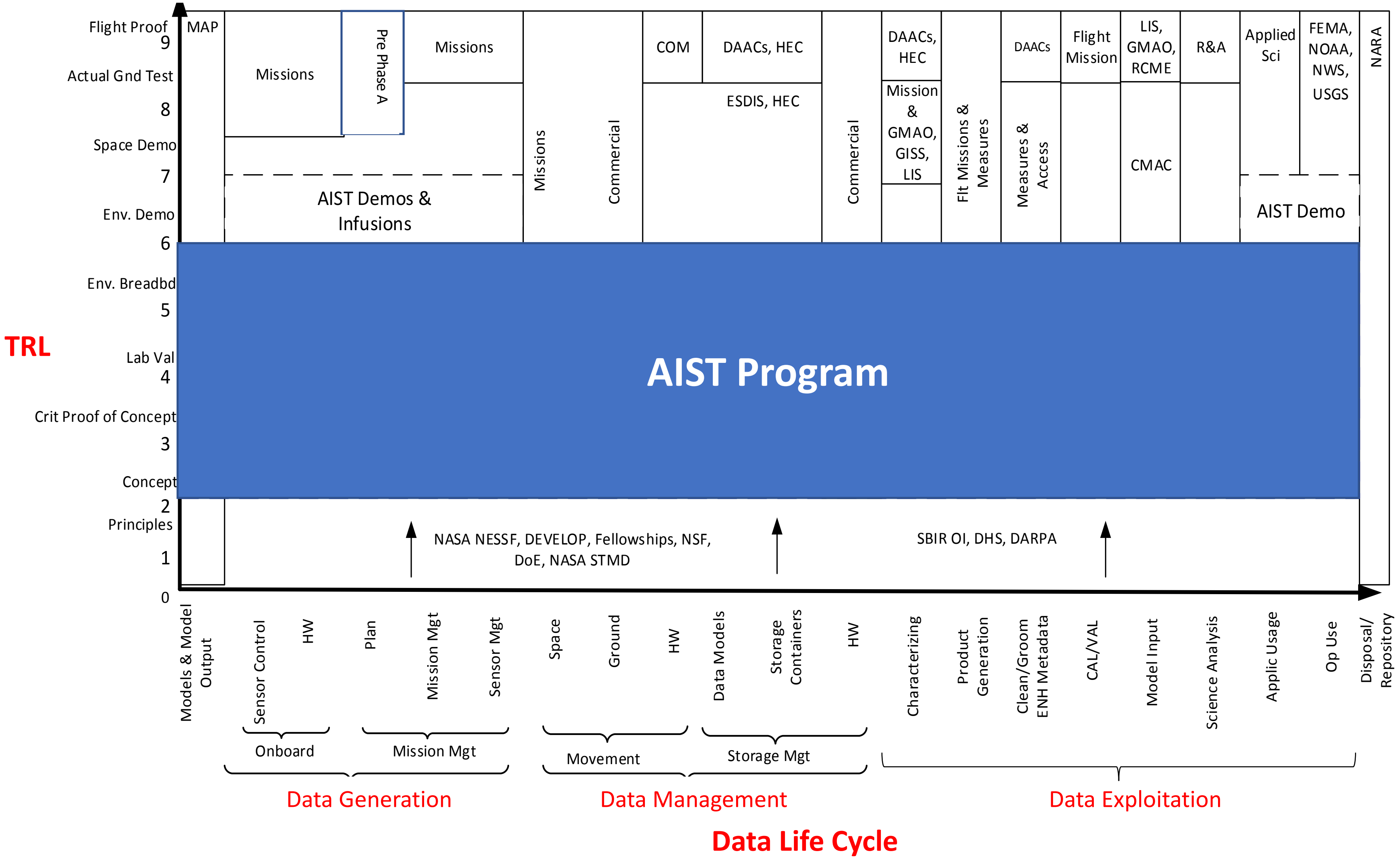
- *Artificial Intelligence (AI)* covers the development of the framework and of the technologies that enable a machine to perceive, reason, plan, act and learn both rationally and humanly.
- *Machine Learning (ML)* covers the sub-field of AI dealing with a machine capable of learning rationally and humanly.
- *Deep Learning (DL)* is a sub-field of ML dealing with very large Artificial Neural Networks including larger numbers of layers and of neurons, trained with massive amounts of data.
- *Machine or Computer Vision (CV)* is the ability to automatically understand any image or video based on visual elements and patterns.

- *Autonomy* is the ability of a system to achieve goals while operating independently of external control.
- *Automation* is the automatically-controlled operation of a system by mechanical or electronic devices that take the place of human labor – *Merriam-Webster*



From Terry Fong "Autonomy NASA Capability Leadership Team (CLT) Internal NASA Presentation", Aug. 2018

AIST Program Scope



AIST Thrusts: NOS and ACF

- **New Observing Strategies (NOS)**

  Optimize measurement acquisition using many diverse observing capabilities, collaborating across multiple dimensions and creating a unified architecture

  - Using Distributed Spacecraft Missions (DSM) or SensorWebs at various vantage points
  - In response to Decadal Survey mission design needs, forecast or science model-driven, or event-driven
  - Using NASA- as well as non-NASA data sources or relevant services

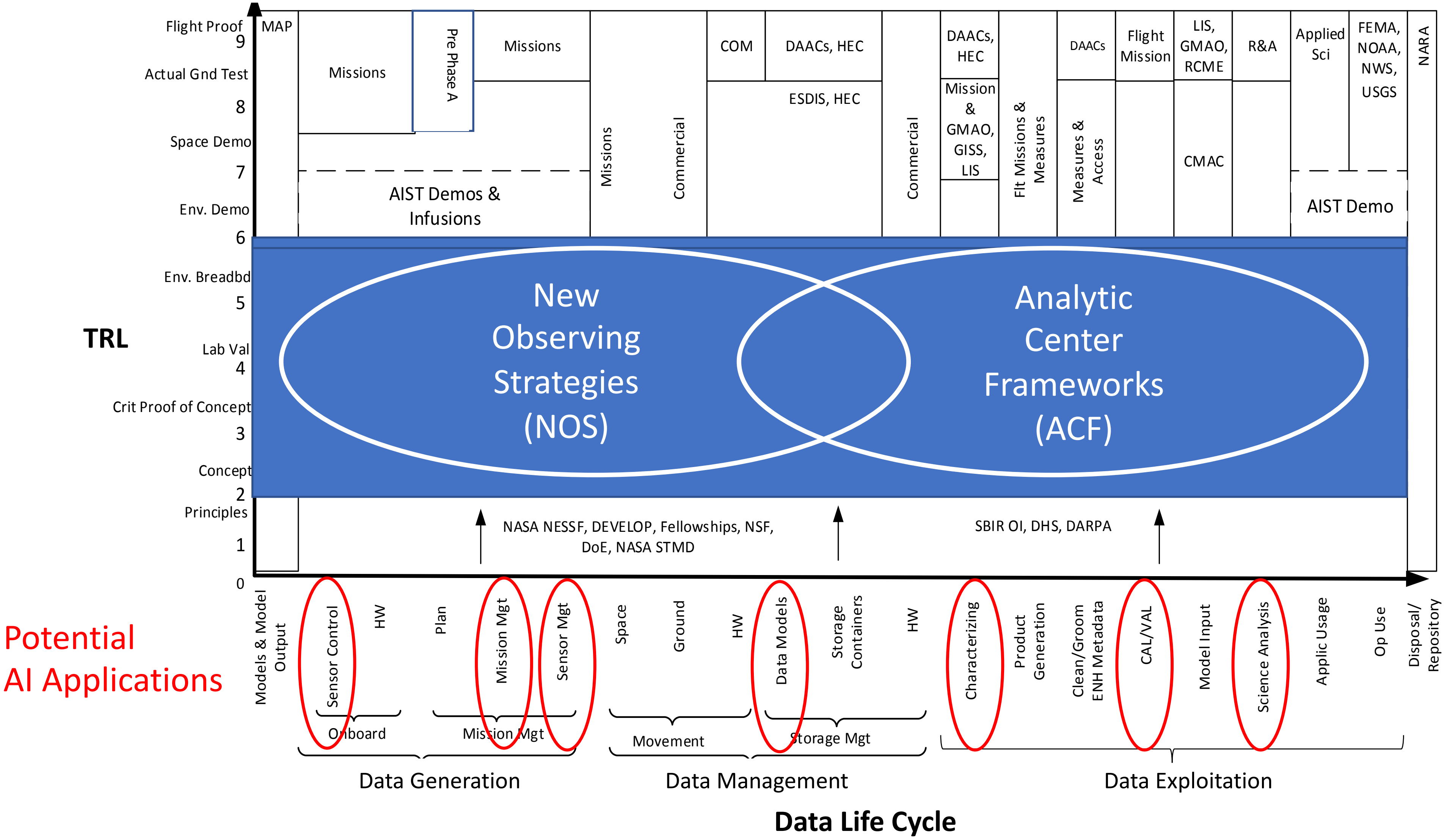
- **Analytic Center Frameworks (ACF)**

  Enhance and enable focused Science investigations by facilitating access, integration and understanding of disparate datasets using pioneering visualization and analytics tools as well as relevant computing environments

  - Allowing flexibility/tailoring configurations for Science investigators to choose among a large variety of datasets & tools
  - Reducing repetitive work in data access and pre-processing, e.g., developing reusable components

- **More generally, address general "Science-Data Intelligence":**
  - "Starting with Science, Ending with Science"
  - Extracting knowledge and information from Science data to make "Science decisions"

AIST Program Scope



AI for Earth Science Applications

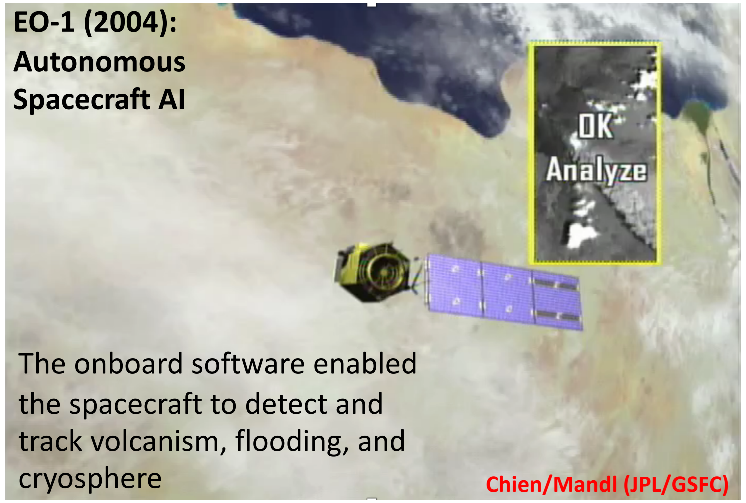
Two Main Areas

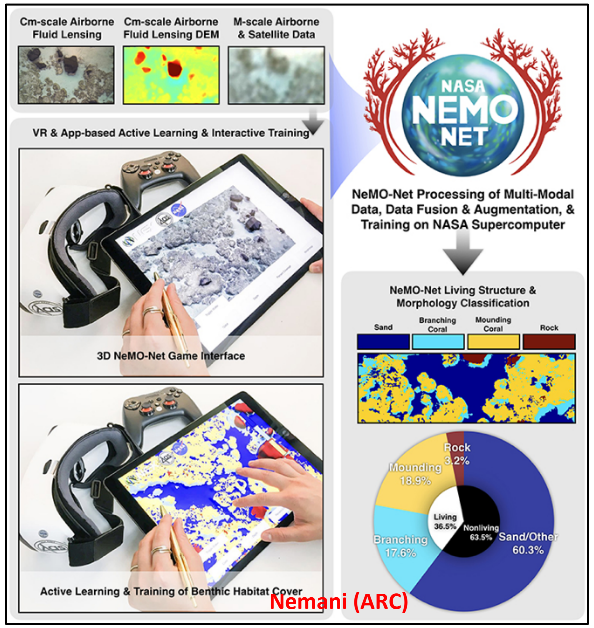
- **Improved Agile Observation Coordination and Mission Operations (Onboard or on the Ground)**
  - At the edge data analysis
  - Semi-autonomy and autonomy for decision making
  - Anomaly and fault detection
  - Engineering Support for large constellations
  - Advanced Interoperability

Technologies: Smart Sensors, Planning & Scheduling, Intelligent Agents, Cognitive and Knowledge-Based Systems, Reasoning, …

- **Science Advancement**
  - Multi-source data integration
  - Big data analytics: discover correlations in large amounts of data
  - Improvements and support to forecasting and science modeling and data assimilation

Technologies: Machine Learning/Deep Learning, Intelligent Search, Computer Vision, Data Fusion, Interactive Visualization & Analytics, Natural Language, …





Some Current or Recent Applications of AI to NASA Challenges

ONBOARD DATA ANALYSIS AND DECISION MAKING and MISSIONS OPERATIONS:

- "Onboard Analysis of Dust Devils on Mars" (Estlin et al)
- "ASPEN: Automated Scheduling and Planning Environment" (Chien)
- "Eagle Eye, specialized ASPEN model for steerable 2D framing instruments" (Knight)
- "Volcano, Fire and Flood Monitoring SensorWebs" (Chien and Mandl)
- "Distributed Spacecraft Autonomy (DSA)" (Micire)
- "Distributed and Multi-Agent Systems" (Frank)
- "Machine Learning and Instrument Autonomy" (Mandrake et al)
- "Supervised-Machine Learning for Intelligent Collision Avoidance Decision-Making and Sensor Tasking" (Mashiku and Memarsadeghi)
- "Autonomous & Tele-Robotics Satellite Servicing" (Roberts)
- "System Engineer's Virtual Assistant (SEVA)" (Coronado)
- "Discovery and Systems Health (DaSH)" (Bajwa)
- "Collaborative and Assistant Systems" (Papasin)
- "Fault Detection, Diagnosis, and Prognostics" (Oza)

SCIENCE DATA ANALYSIS/DATA ANALYTICS- ON THE GROUND:

- "Wildfires and Solar Events Detection using Deep Learning" (MacKinnon, Ichoku and Pulkinnen)
- "Registration, Super-Resolution and Data Assimilation using Machine Learning," (Halem, Pelissier, Ames)
- "Multi-Temporal Analysis of Remote Sensing Data" (Pinzon and Tucker)
- "Data Analytics and Machine Learning for NASA Heliophysics Data" (Thompson)
- "Big Data Technologies for Radio Astronomy" (Jones)

AI in ESTO Advanced Information Systems Technology (AIST) Projects

- **AI for Observation Simulation Synthesis Experiments (OSSEs) and for Mission Design**
  - A Mission Planning Tool for Next Generation Remote Sensing of Snow (Forman/AIST-16)
    As part of a new simulation tool that will help identify the best combination of satellite sensors to detect snow and measure its water content from space, Machine Learning maps model states into observation space; in particular, Machine Learning has been used to predict C-band SAR backscatter over snow-covered terrain in Western Colorado using a support vector machine (SVM). Backscatter coefficients were obtained via supervised training using observations from the European Space Agency's Sentinel-1A and Sentinel-1B sensors.
  - Trade-space Analysis Tool for Constellations Using Machine Learning (TAT-C ML) (Verville & Grogan/AIST-16)
    TAT-C is a systems architecture analysis platform for pre-phase A Earth science (ES) constellation missions. It allows users to specify high-level mission objectives and constraints and efficiently evaluate large trade spaces of alternative architectures varying the number of satellites, orbital geometries, instruments, and ground processing networks. Outputs characterize various mission characteristics and provide relative evaluations of cost and risk. Machine Learning evolutionary algorithms are used for fast traversal of this large trade space using Adaptive Operator Selection (AOS) and Knowledge-driven Optimization (KDO) working with a Knowledge Base populated with information from historical ES missions.

- **AI for Time Series and for Science Models**
  - Advanced Phenology Information System (APIS) (Morisette/AIST-16)
    Ecological processes and uncertainty are evaluated by fitting a Bayesian hierarchical model to annually oscillating time series of vegetation indices, with the R package "greta", which utilizes TensorFlow and the TensorFlow Probability module. This enables to make inference not only on site- or year-specific patterns in the historical record, but also on the drivers of phenology, including proper estimates of prediction uncertainty. This model allows to make good predictions for years for which there is very limited data.
  - NASA Evolutionary Programming Analytic Center (NEPAC) (Moisan/AIST-18)
    NEPAC's main objective is to demonstrate a Machine Learning application, called Genetic Programming of Coupled Ordinary Differential Equations (GPCODE), that uses a combination of Genetic Programming (GP) and Genetic Algorithms (GA) to automatically generate optimized algorithms for satellite observations and coupled system of equations for ecosystem models. NEPAC will initially focus on evolving new ocean chlorophyll algorithms using an expanded set of performance metrics and a regression technique, called Maximum Probability Regression (MPR), that requires estimates of the optimization data set's error, variance and co-variances.
  - Canopy Condition to Continental Scale Biodiversity Forecasts (Swenson/AIST-18)
    The goal is to characterize canopy condition from various spatio-temporal remote sensing products (including drought indices and habitat structure) to predict the supply of mast resources to herbivores (and threatened species) and visualize canopy condition. Hyperspectral bands are analyzed to identify relationships between hyperspectral imagery, canopy traits, such as sugar to starch, lignin to non-structural carbohydrates, and overall mast production. This will be done using a Generalized Joint Attribution Model (GJAM) and machine learning algorithms such as a support vector machine (SVM) as well as classic model-based approaches.

AI in ESTO Advanced Information Systems Technology (AIST) Projects

- **AI for Pattern and Information Extraction**
  - Computer-Aided Discovery and Algorithmic Synthesis for Spatio-Temporal Phenomena in InSAR (Pankratius/AIST-16)
    The project goal was to facilitate the discovery of surface deformation phenomena in space and time in InSAR/UAVSAR data. Machine Learning, specifically neural networks, was used to identify which parts of InSAR interferograms are primarily caused by tropospheric effects versus real surface deformations. Because of sparse training sets, representative InSAR data is perturbed and used to simulate data where it is missing, thus augmenting the training dataset. Information from the domain knowledge, rules of geophysics and atmospheric science are used as a way to overcome the sparsity problem.

  - Autonomous Moisture Continuum Sensing Network (Entekhabi & Moghaddam/AIST-16)
    Soil moisture is important for understanding hydrologic processes by monitoring the flow and distribution of water between land and atmosphere. A distributed, adaptive ground sensor network improves observations while reducing energy consumption to extend field deployment lifetimes. Embedded Machine learning decides when and where to sample in order to optimize information gain and energy usage. Alternative adaptive sampling strategies have been evaluated for performance i.e., maximizing information gain) vs energy use. Autoregressive Machine Learning was demonstrated to have superior performance. The project is currently collaborating with the CYGNSS mission for cal/val activities.

  - Supporting Shellfish Aquaculture in the Chesapeake Bay using AI for Water Quality (Schollaert-Uz/AIST-18)
    Provide access to reliable information on a variety of environmental factors, not currently available at optimal scales in space and times, by using various data (sats and others) and AI for Pattern Recognition.
    Increasing use of machine learning to address geoscience questions offers the potential to detect contextual cues from large unstructured datasets to extract patterns such as locations of poor water quality. This project will use Machine Learning on input data such as temperature and bacterial count to yield output such as a poor water quality indicator. A preliminary unsupervised cluster analysis will determine the most promising parameters that will be used as input features to a neural network, e.g., a Convolutional Neural Network, to perform image semantic segmentation and classify each pixel into a fixed set of categories. Efficient implementation of unsupervised data clustering, data fusion, and interpolation algorithms will be investigated and integrated in the final approach.

  - Mining Chained Modules in Analytics Center Frameworks (Zhang/AIST-18)
    The project's goal is to build a workflow system, as a building block for Analytic Center Frameworks, capable of recommending to Earth Scientists multiple software modules, already chained together as a workflow. The tool will leverage Jupyter Notebooks to mine software module usage history, and to develop algorithms by extracting reusable chains of software modules and then will develop an intelligent service that provides for personalized recommendations.

AI in ESTO Advanced Information Systems Technology (AIST) Projects

- **AI for Image Processing and for Data Fusion**
  - Software Workflows and Tools for Integrating Remote Sensing and Organismal Occurrence Data Streams to Assess and Monitor Biodiversity Change (Jetz/AIST-16)

    When considering large numbers of biodiversity records, the most efficient way to retrieve values is to minimize the number of scene calls and maximize useful data outputs from each call. To optimize efficiency, clustering (i.e., spatial and temporal aggregations) is implemented in which input values are grouped to optimize efficiency; input values are grouped in three dimensions (latitude, longitude, and time) into clumps that fall into the same scenes and reduce the number of scene calls. Different clustering techniques are applied, depending on the spatiotemporal resolution of the environmental product. Each 'cluster' additionally serve as the unit of parallelization of processing.

  - NeMO-Net - The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment (Chirayath/AIST-16)

    The project goal is to assess global present and past dynamics of coral reef systems. An invariant algorithm was created that combines Convolutional Neural Networks (CNN) and traditional Machine Learning techniques (e.g., K-Nearest Neighbors) to predict shallow marine benthic classes to a high degree of accuracy. The deep neural networks were trained using a citizen science app that allows people to label images. The algorithm was trained and tested on WorldView 2 imagery, and then used directly to successfully process Planet imagery. By using transfer learning and domain adaptation, NeMO-Net demonstrates data fusion of regional FluidCam (mm, cm-scale) airborne remote sensing with global low-resolution (m, km-scale) airborne and spaceborne imagery to reduce classification errors up to 80% over regional scales.

- **AI for Quantum Computing**
  - Framework for Mining and Analysis of Petabyte-size Time-series on the NASA Earth Exchange (NEX) (Michaelis & Nemani/AIST-16)

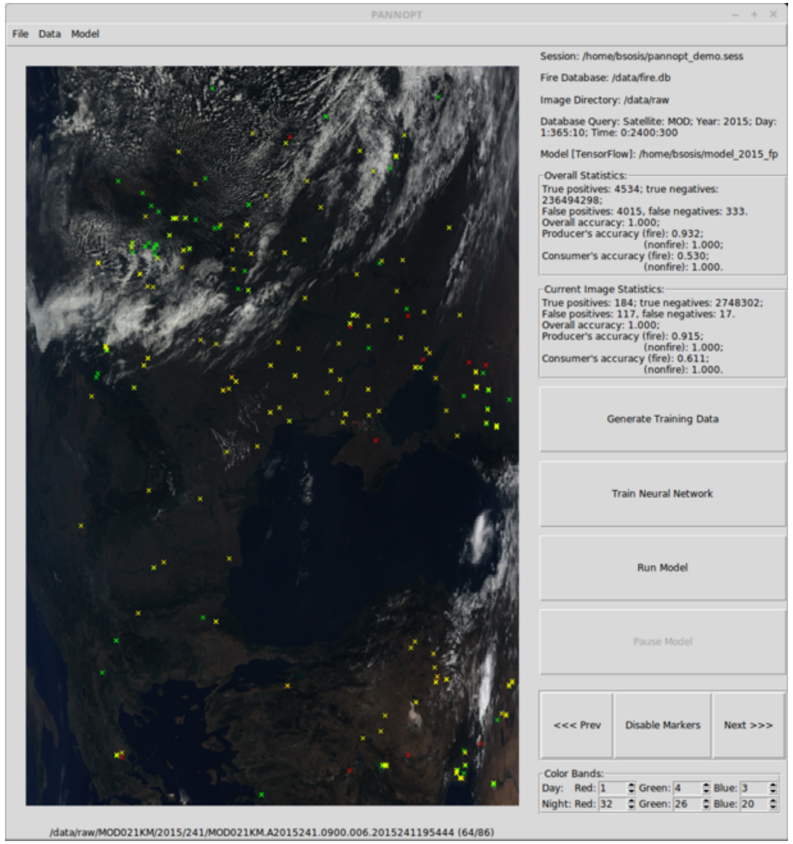
    The project goal is to create a capability for fast and efficient mining of time-series data from NASA's satellite-based observations, model output, and other derived datasets. As part of this project, a quantum assisted generative adversarial network (GAN) has been implemented for both quantum assisted transformation/compression (QAT) and for machine learning based time-series analytics. The method has been implemented on the D-Wave 2000Q, using around 1500 (out of available 2048) qubits.

  - An Assessment of Hybrid Quantum Annealing Approaches for Inferring and Assimilating Satellite Surface Flux Data into Global Land Surface Models (Halem/AIST-16)

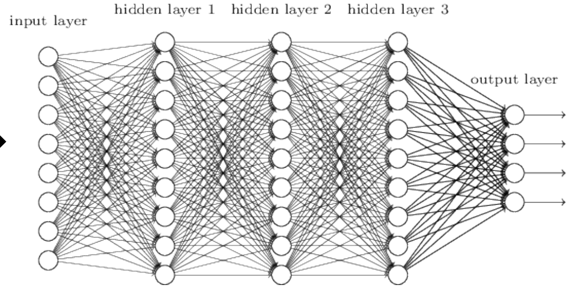
    The main goal of this project is to demonstrate the scope of Hybrid Quantum Annealing algorithmic research to support NASA Earth science on the next generation of D-Wave architectures. As part of this project, Machine Learning was investigated for several applications including the use of Recurrent Neural Networks (RNNs) with Long Short Term Memory (LSTM) models for predicting CO2 fluxes, investigating how machine learning can be applied to mapping global carbon flux with Fluxnet data, generating global continuous solar-induced chlorophyll fluorescence (SIF) based on OCO-2 data and Neural Networks. and image registration using both Discrete Cosine Transform and Botzmann Machine.

CNN Example (MacKinnon/GSFC)

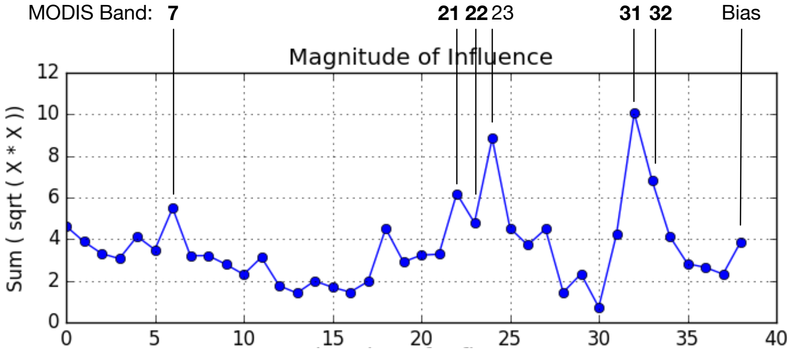
Wildfire Classification in MODIS Data Using Deep Learning



Training GUI developed in Python to automate training set generation, model running, and data visualization



Combined approach of training a Deep NN on the ground using high-performance GPUs with a highly optimized inference system running on a flight-proven embedded processor => 99.2 % correct classification rate for the detection of wildfires in MODIS data. Tested onboard an Unmanned Airborne System (UAS)



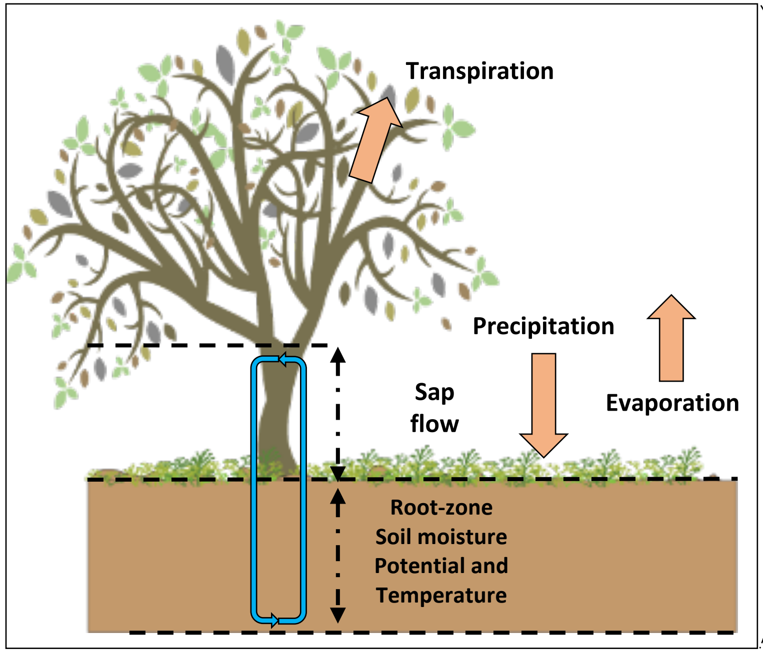
Validation Results (samples not in training set):

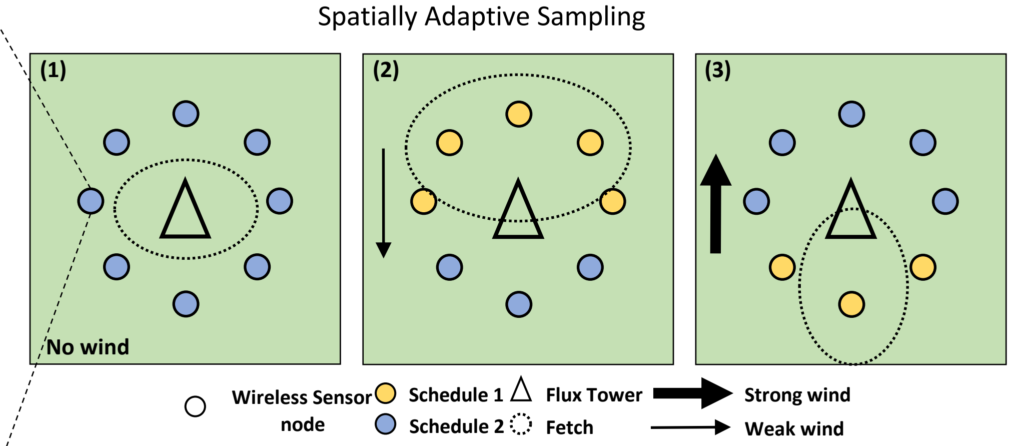
- A preliminary look at the NN weights after training indicate that the MODIS 7, 21, 22, 23, 31, and 32 bands were the most influential for detecting fires.
- Note the MODIS active fire detection algorithm uses bands: 1, 2, 7, 21, 22, 31, and 32.

AIST-16/Entekhabi & Moghaddam (MIT & USC) –

Autonomous Moisture Continuum Sensing Network

Soil moisture is important for understanding hydrologic processes by monitoring the flow and distribution of water between land and atmosphere. A distributed, adaptive sensor network improves observations while reducing energy consumption to extend field deployment lifetimes .





Distributed wireless sensor network measures soil moisture, sap flow, and winds

Embedded Machine learning decides when and where to sample in order to optimize information gain and energy usage.

Evaluated alternative adaptive sampling strategies for performance (information) vs energy use.

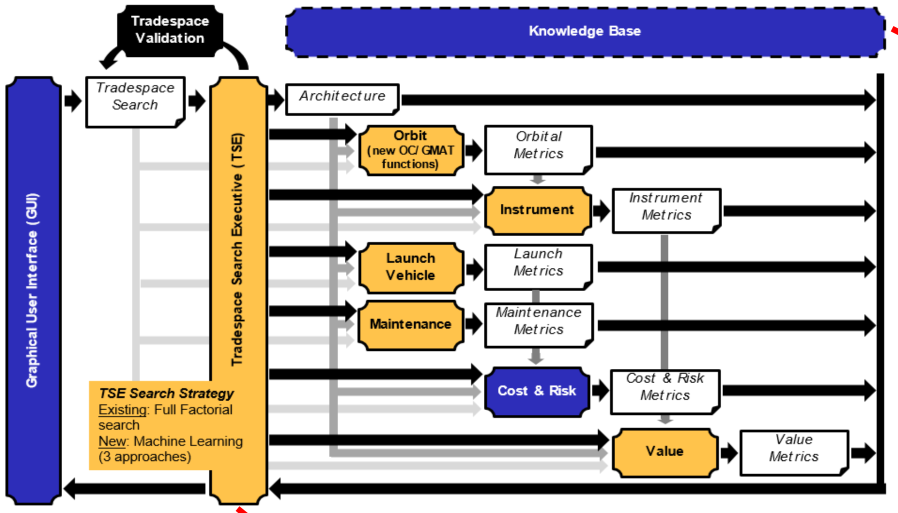
- ✓ Information Gain vs. Energy Consumption optimization → present as Pareto Fronts
- ✓ An autoregressive ML will have superior performance $\theta(t) = f\big(\theta(t-1)\big) + g\big(X(t)\big)$
- ✓ Simple Policies can achieve superior RMSE performance with less energy consumption

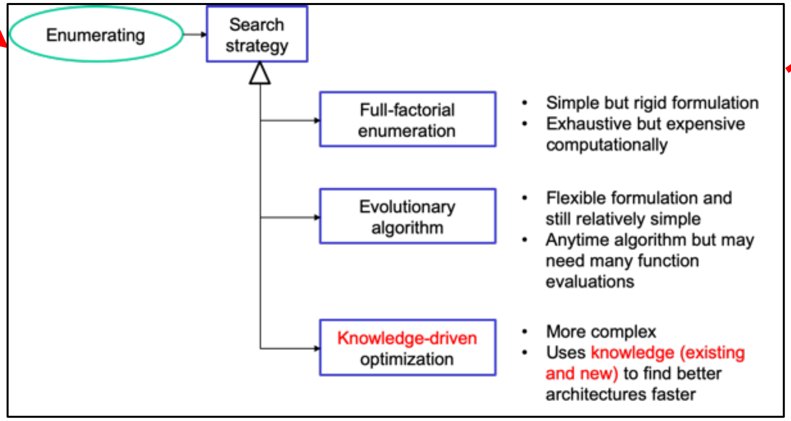
- SoilSCAPE Plan → Satellites Cal/Val
  - SMAP Cal/Val: Deployed 1 site at the Cary Institute of Ecosystem Studies (Millbrook, NY)
  - SoilSCAPE team (via. Co-I Moghaddam) collaborating with CYGNSS to provide *in situ* soil moisture for cal/val activities
  - Established a cal/val infrastructure for NiSAR

SoilSCAPE installation for CYGNSS Cal/Val

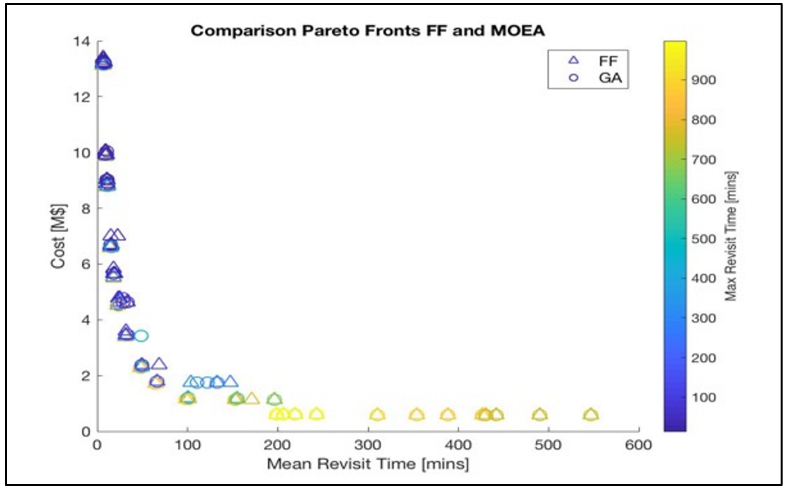
AIST-14 & -16/Grogan (Stevens) –
Trade-space Analysis Tool for Constellations (TAT-C)

TAT-C is a systems architecture analysis platform for pre-phase A Earth science (ES) constellation missions. It allows users to specify high-level mission objectives and constraints and efficiently evaluate large trade spaces of alternative architectures varying the number of satellites, orbital geometries, instruments, and ground processing networks. Outputs characterize various mission characteristics and provide relative evaluations of cost and risk. Machine Learning evolutionary algorithms are used for fast traversal of this large trade space using Adaptive Operator Selection (AOS) and Knowledge-driven Optimization (KDO) working with a Knowledge Base populated with information from historical ES missions.





Evolutionary algorithm finds a very similar Pareto front with an order of magnitude fewer function evaluations



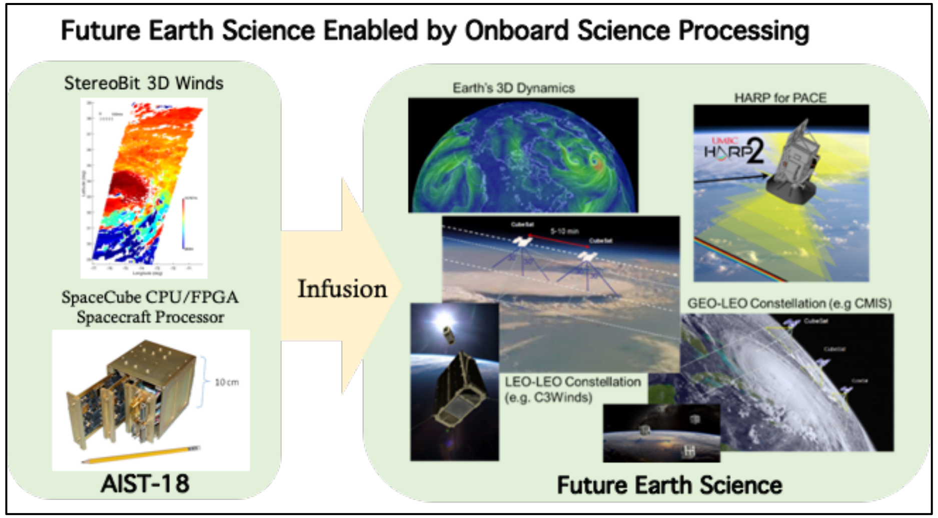
Evolutionary algorithm enables searching much larger and richer design spaces with hybrid architectures combining satellites at various altitudes and inclinations.

Delta heterogeneous Walker	
Decision	**Options**
# satellites	[1, 2, 3, 4, 6, 8, 10, 12]
# planes	[1, 2, 3, 4, 6, 8, 12]
Altitude	[400:100:800] km
Inclination	[0°,10°,20°,30° ISS, 90°, SSO]

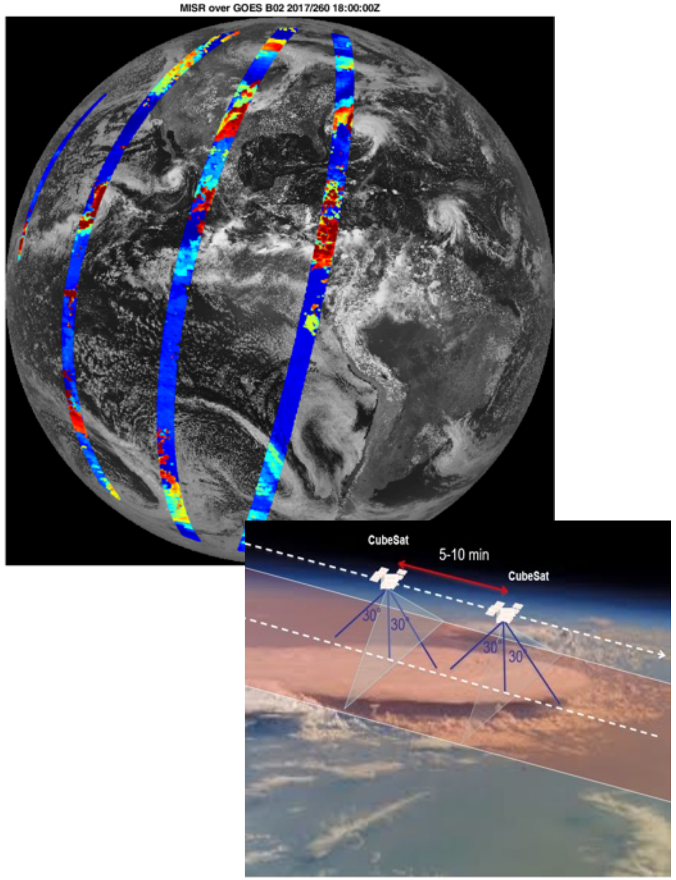
Note: Previous PI's: Le Moigne & Verville @GSFC

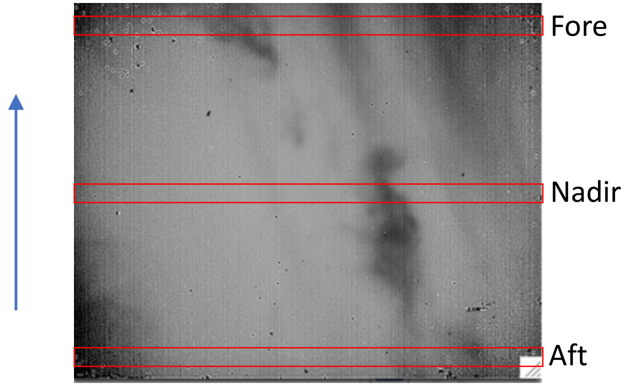
AIST-18/Carr (Carr Astronautics) –
StereoBit: Advanced Onboard Science Data Processing to Enable Future Earth Science

This investigation will demonstrate higher-level onboard science data processing for more intelligent SmallSats and CubeSats to enable future Earth science missions and Earth observing constellations. Low-cost SmallSat architectures generally suffer from downlink bottlenecks and often result in lower data acquisitions per orbit. This project targets an objective relevant to the 2017-2027 Earth Sciences Decadal Survey - atmospheric dynamics with 3D stereo tracking of cloud moisture features using a Structure from Motion (SfM) technique called StereoBit that can be implemented onboard. This will lead to the development of a testbed to validate intelligent onboard systems.



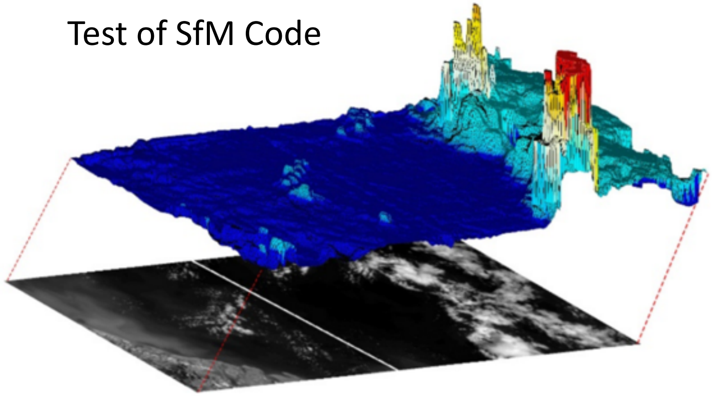
SfM method from OpenCV implemented on SpaceCube 2.0 and flying on RRM3 using the Compact Thermal Imager (CTI)





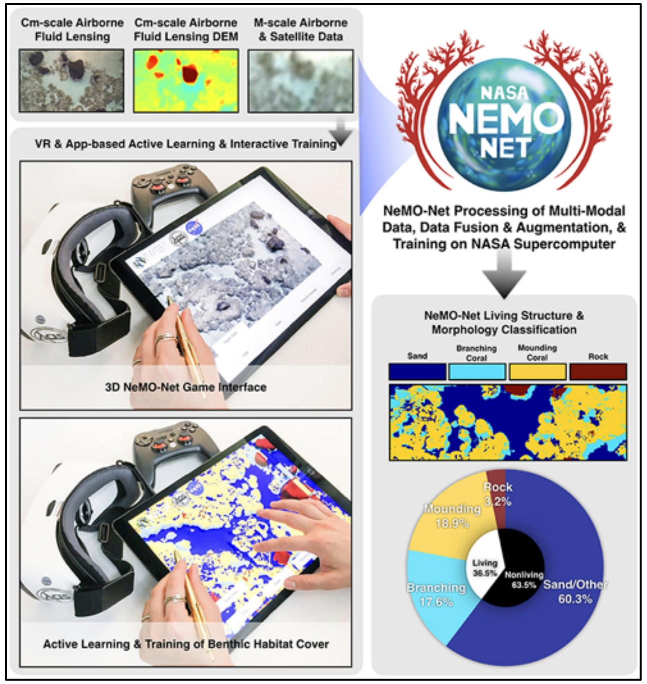
Early CTI Cloud Picture

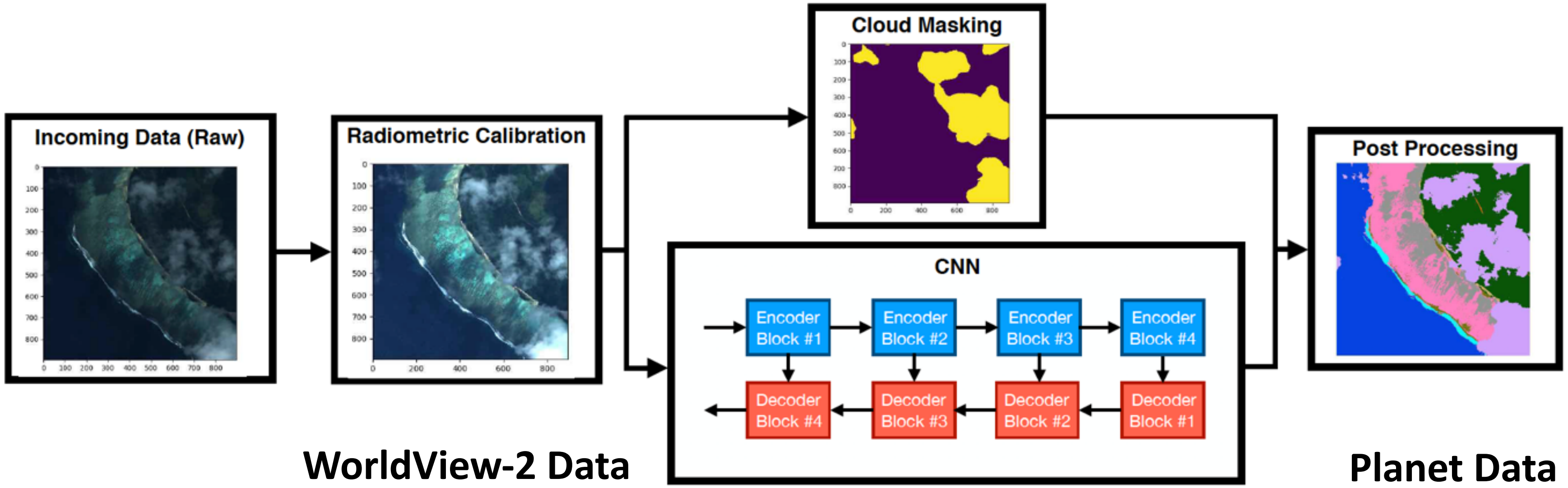
Test of SfM Code



AIST-16/Chirayath (ARC) –
NeMO-Net - The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment

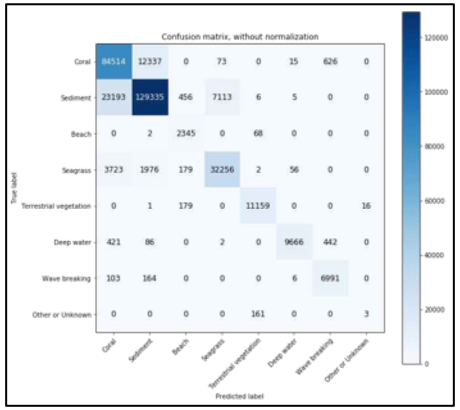
NeMO-Net deploys deep convolutional neural networks (CNNs) and active learning techniques to accurately assess the present and past dynamics of coral reef ecosystems through determination of precent living cover and morphology as well as mapping of spatial distribution. Ingests heterogeneous data from airborne and satellite imagery to demonstrate data fusion techniques to resolve temporal, spectral and spatial differences across datasets; and extends predictions over large temporal scales. The deep neural networks were trained using a citizen science app that allows people to label images. The algorithm was trained and tested on WorldView 2 imagery, and then used directly to successfully process Planet imagery.



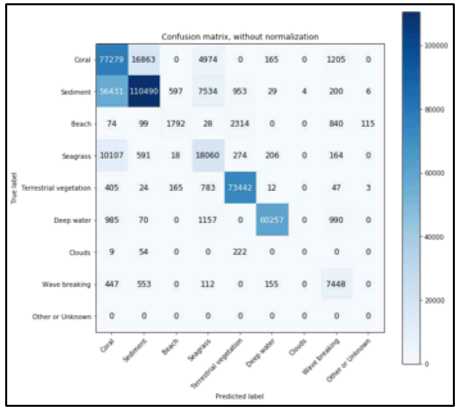


WorldView-2 Data

Planet Data



Total Accuracy: 84.3%
Accuracy amongst coral, sediment, seagrass classes: 83.6%



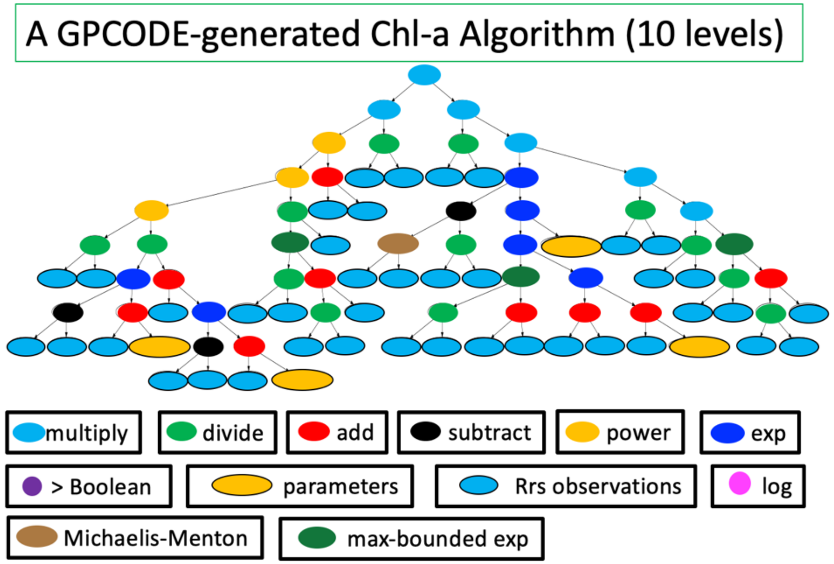
Total Accuracy: 76.0%
Accuracy amongst coral, sediment, seagrass classes: 68.1%

AIST-18 – *Using AI for Facilitating the Development of Novel Science Data Processing Workflows and Algorithms*

Moisan (GSFC)/"NASA Evolutionary Programming Analytic Center (NEPAC)"

Develop a framework that uses genetic programming to generate new chlorophyll-a (chl-a) algorithms with reduced uncertainties and anneal chl-a estimates across multiple ocean (OC) satellite data sets.

User Requirements/variables (e.g., Sea Surface Temperature, Salinity) & Training Datasets
=> Generate Tree of parameters => Optimal Equation(s)

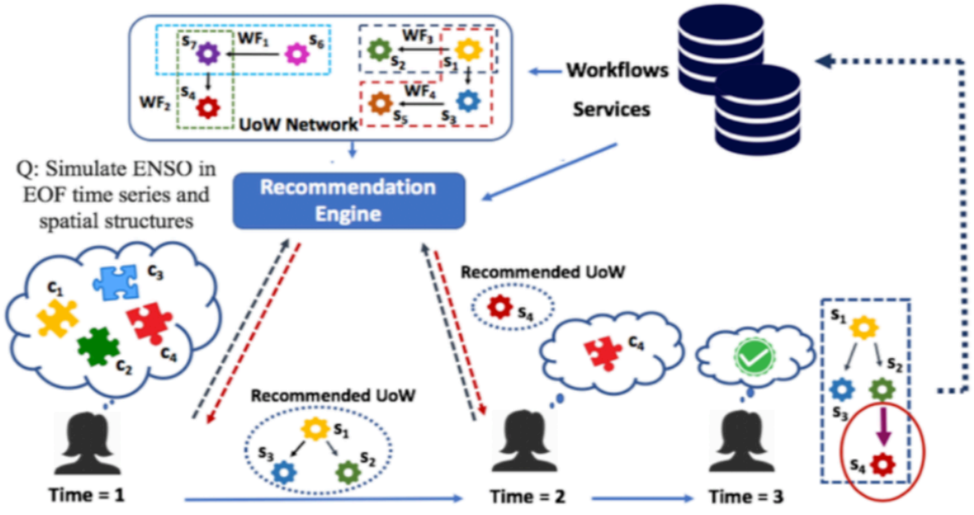


Improved Ocean Color Algorithms and Climate Data Records Using Machine Learning/Genetic Programming

Zhang (CMU)/"Mining Chained Modules in Analytic Center Framework"

Build a workflow tool as a building block for Analytic Center Frameworks that is capable of recommending chained software modules.

Mine software usage history => construct a knowledge network => explore reusable software module chains => Intelligent service for personalized workflow design recommendations

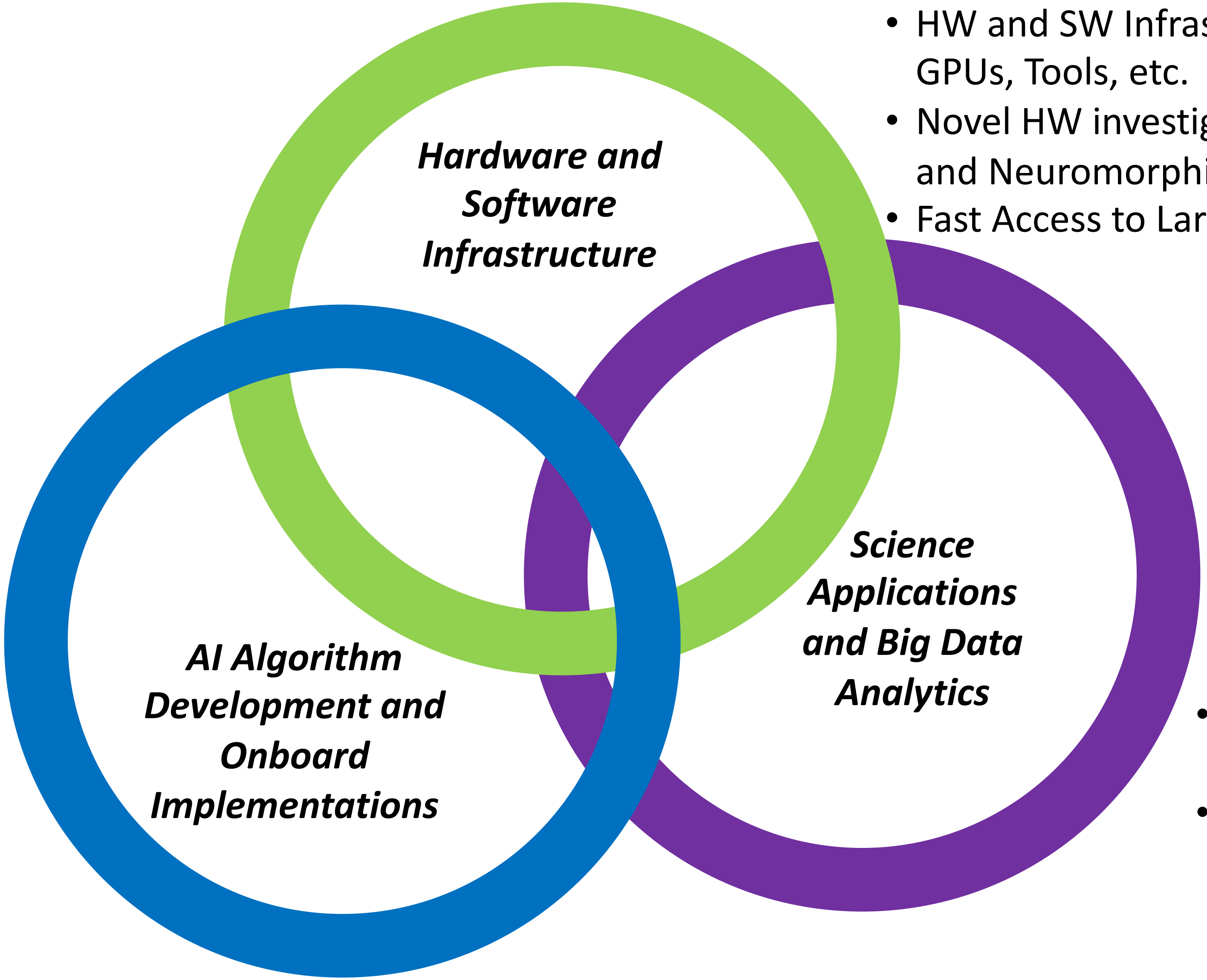


Workflow Recommendation Framework

Future: AI Strategy for NASA Applications

- HW and SW Infrastructure , including GPUs, Tools, etc.
- Novel HW investigation, e.g., Quantum and Neuromorphic Computing
- Fast Access to Large Amounts of Data

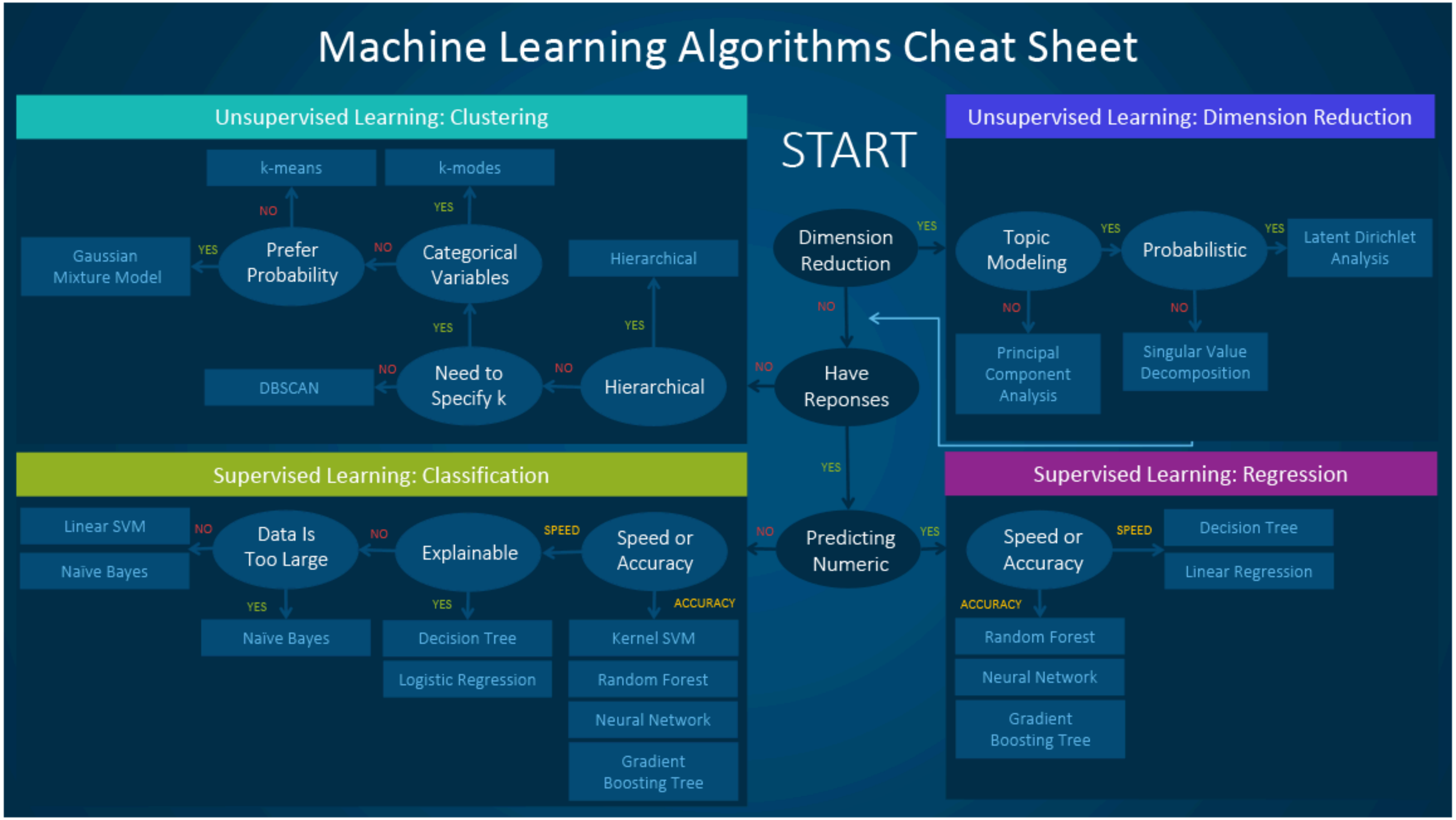
- AI Expertise
- Conceptual Software & Algorithm Development
- At the Edge/Onboard Implementations



- Science Applications and Data Analytics
- Algorithm Relevance and Validation

How to Choose a ML Algorithm

https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/

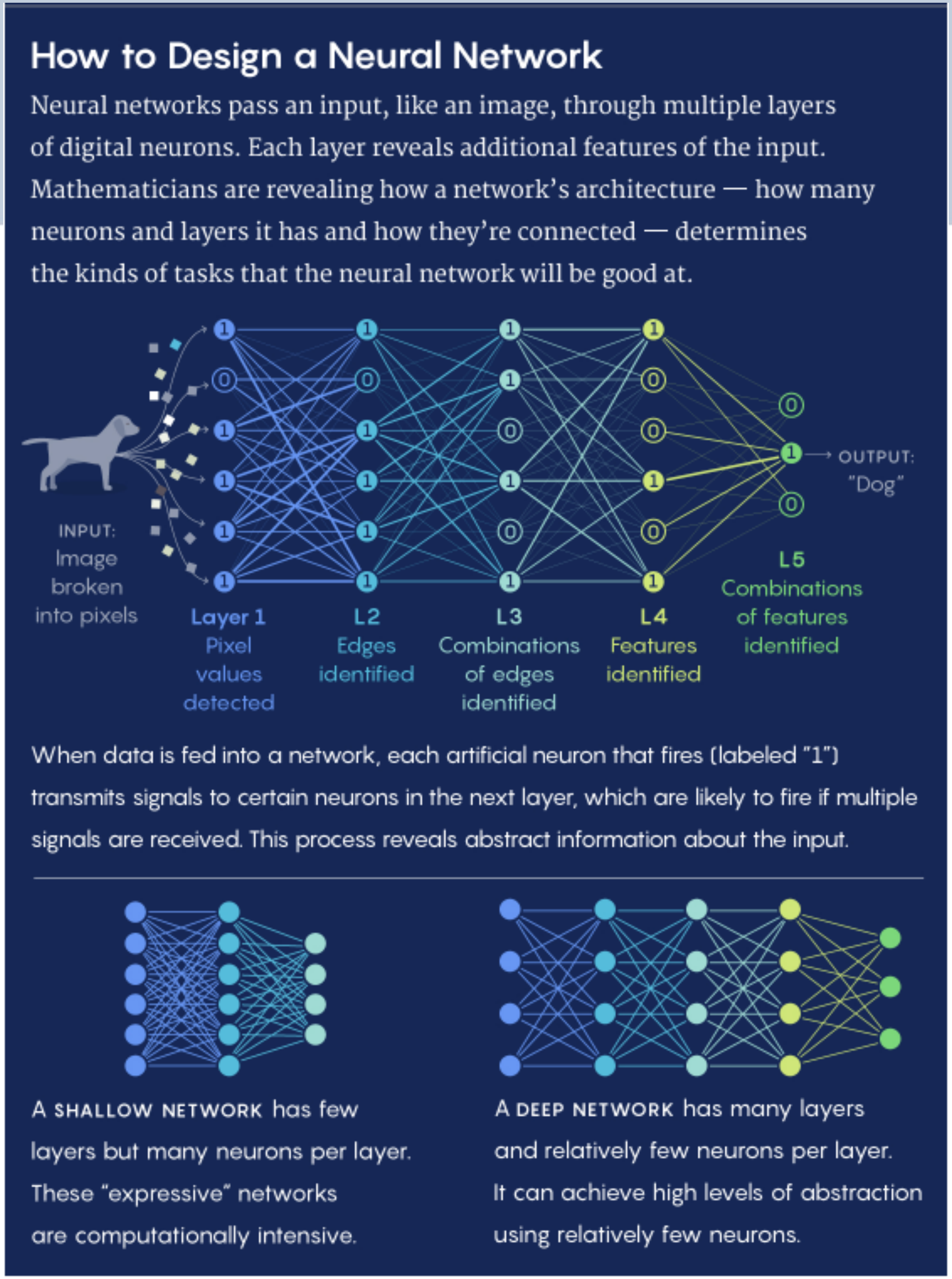


How do you know which neural network architecture will accomplish a task best?

From https://www.quantamagazine.org/foundations-built-for-a-general-theory-of-neural-networks-20190131/

There are some broad rules of thumb:

- *For image-related tasks*, engineers typically use "convolutional" neural networks (CNN), which feature the same pattern of connections between layers repeated over and over.
- *For natural language processing* — like speech recognition, or language generation — engineers have found that "recurrent" neural networks (RNN) seem to work best. In these, neurons can be connected to non-adjacent layers.
- However, engineers largely have to rely on *experimental evidence ("trial and error")*: they run 1,000 different neural networks and simply observe which one gets the job done.
- Rolnick and Tegmark/MIT proved that by *increasing depth and decreasing width, you can perform the same functions with exponentially fewer neurons.* They found that there is power in taking small pieces and combining them at greater levels of abstraction instead of attempting to capture all levels of abstraction at once.
- *BUT,* other researchers have proven that *no amount of depth can compensate for a lack of width*, *e*.g., for sheep recognition, "if none of the layers are thicker than the number of input dimensions, there are certain shapes the function will never be able to create, no matter how many layers you add." (Johnson)



Future - Some ML Technologies To Explore Further

- *Ebert-Uphoff, Samarasinghe and Barnes (2019), "Thoughtfully Using Artificial Intelligence in Earth Science"*
  - *Why Use AI for my application?*
  - *Can I leverage scientific knowledge?*
  - *Can I leverage explainable AI?*
  - *Does my approach generalize?*
  - *Are my results reproducible?*
  - *Do I generate new scientific insights?*
- Areas of research
  - **Examine and characterize various ML technologies for various applications and Compare performance with classical approaches**
    - Benchmark datasets and applications?
  - **Integrate physical models into machine learning algorithms:**
    - Bring a priori understanding of specific scientific and engineering problems into the decision-making process
    - Improve performance, but to significantly increase the interpretability of the model by relating its parameters and outputs to well-understood physical quantities
  - **Explainable AI** techniques as they apply to NN, e.g.:
    - Explore combined use of knowledge-base and data mining techniques applied to analyzing NN processes
    - Visualization of processes using Augmented Reality/Virtual Reality (AR/VR)
  - **Transfer Learning** to generalize, reuse and/or adapt previously developed algorithms and minimize the amount of training data required
  - **Applications to scientific discovery**
    - Data/Image interpretation and mining from distributed archives
    - Non-obvious information discovery through fusion of multiple datasets
    - Uncertainty quantification
- **Other Applications to pursue**
  - Personalized Expert Assistant, e.g., Document/literature search and connect relevant/personalized information
  - Onboard computing for data reduction, intelligent data/image interpretation, data fusion, etc.

Any Questions?

Earth Science Technology Office