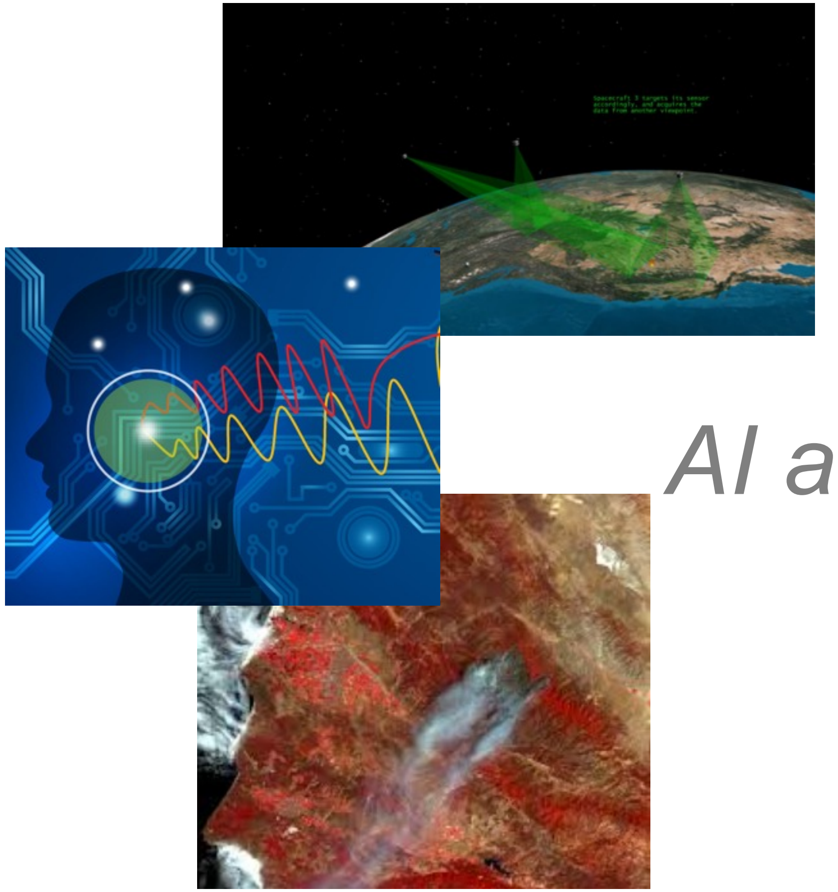
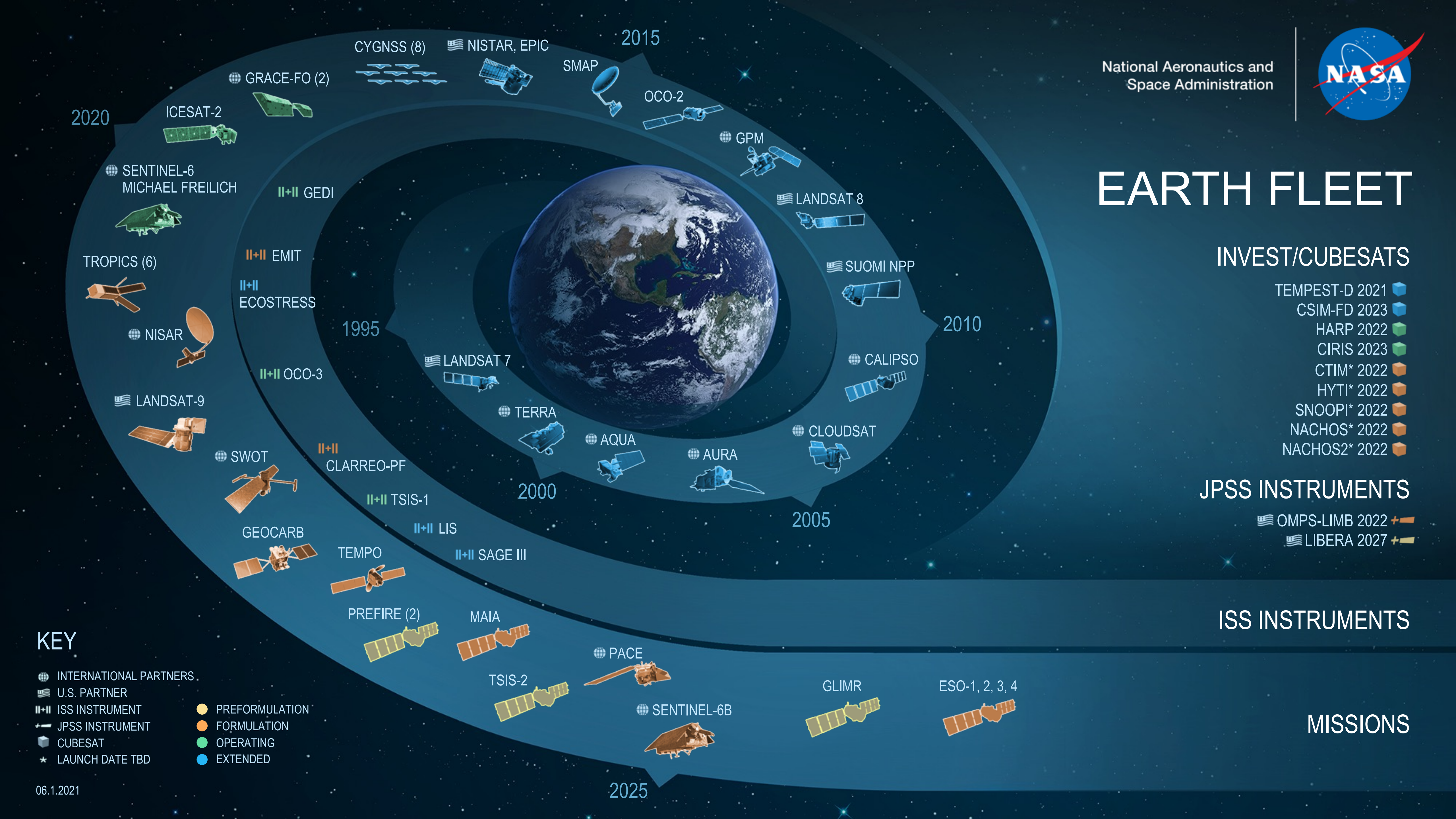
Artificial Intelligence and Machine Learning for Earth Science



ISU Alumni Conference 2021
AI and Big Data Applications to Space

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

July 31, 2021



National Aeronautics and Space Administration
NASA
EARTH FLEET
INVEST/CUBESATS
TEMPEST-D 2021
CSIM-FD 2023
HARP 2022
CIRIS 2023
CTIM* 2022
HYTI* 2022
SNOOPI* 2022
NACHOS* 2022
NACHOS2* 2022
JPSS INSTRUMENTS
OMPS-LIMB 2022
LIBERA 2027
ISS INSTRUMENTS
MISSIONS
2015
2020
2010
1995
2000
2005
2025
CYGNSS (8)
NISTAR, EPIC
SMAP
OCO-2
GPM
LANDSAT 8
SUOMI NPP
CALIPSO
CLOUDSAT
AURA
AQUA
TERRA
LANDSAT 7
GRACE-FO (2)
ICESAT-2
SENTINEL-6 MICHAEL FREILICH
TROPICS (6)
NISAR
LANDSAT-9
SWOT
GEOCARB
TEMPO
PREFIRE (2)
MAIA
TSIS-2
PACE
SENTINEL-6B
GLIMR
ESO-1, 2, 3, 4
GEDI
EMIT
ECOSTRESS
OCO-3
CLARREO-PF
TSIS-1
LIS
SAGE III
KEY
INTERNATIONAL PARTNERS
U.S. PARTNER
ISS INSTRUMENT
JPSS INSTRUMENT
CUBESAT
LAUNCH DATE TBD
PREFORMULATION
FORMULATION
OPERATING
EXTENDED
06.1.2021

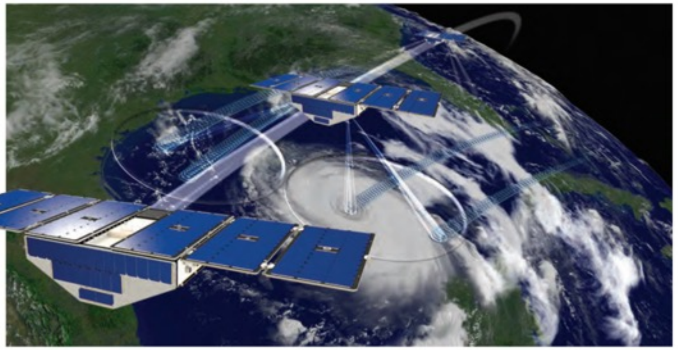
Advanced Information Systems Technology
AIST Objectives

AIST Technology Program is funding software and information systems, as well as novel computer science technologies expected to be needed by Earth Science Missions and Programs in the 5-10-year timeframe.
AIST currently has three thrusts:

- **New Observing Strategies (NOS):** New observation measurements and new observing systems design and operations through intelligent, timely, dynamic, and coordinated distributed sensing
- **Analytic Collaborative Frameworks (ACF):** Agile science investigations that fully utilize the large amount of diverse observations using advanced analytic tools, visualizations, and computing environments, and that interact seamlessly with relevant observing systems;
- **Earth System Digital Twins (ESDT):** Development of integrated Earth Science frameworks that mirror the Earth with state-of-the-art models (Earth system models and others), timely and relevant observations, and analytic tools.

NOS for Optimizing Measurements Design & Dynamically Capturing full Science Events

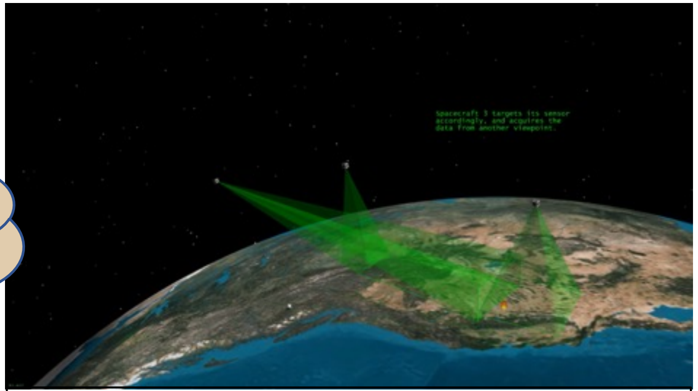
CYGNSS microsatellite observatories in orbit. Image credit: Southwest Research Institute



Distributed Spacecraft Mission (DSM): mission involving multiple spacecraft to achieve one or more common goals.

Provide complete picture of physical processes or natural phenomena

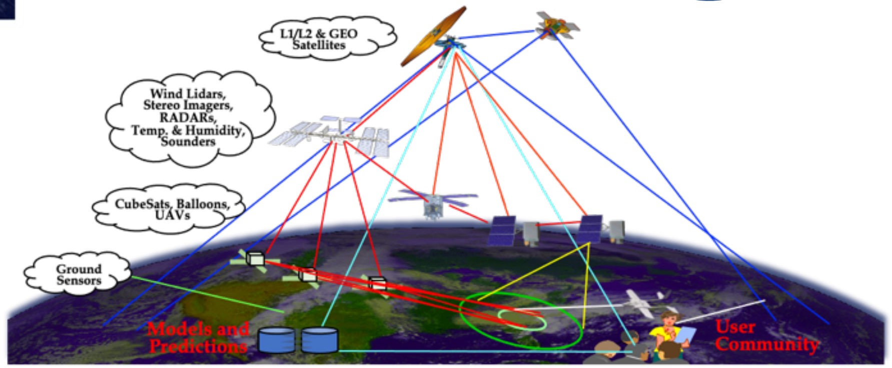
Increased understanding and predictability of dynamic events on Earth.



A special case of DSM is **an Intelligent and Collaborative Constellation (ICC)** which involves the combination of:

- Real-time data understanding
- Situational awareness
- Problem solving;
- Planning and learning from experience
- Communications & cooperation between several S/C

Multiple collaborative nodes from multiple organizations (NASA, OGAs, Industry, Academia, International) from multiple vantage points and in multiple dimensions (spatial, spectral, temporal, radiometric)



A SensorWeb is a distributed system of ***sensing nodes*** (space, air or ground) that are interconnected by a ***communications fabric*** and that functions as a single, highly coordinated, virtual instrument.

Actively acquire data in coordination with other sensors, models in response to measurement needs and/or science events

OBJECTIVES:

1. **Design and develop New Observing Concepts:**
   - From Decadal Survey or Model; **Various size spacecraft;** Systems of systems **(*Internet-of-Space*); Various organizations**
   - **Perform trades** on sensor number/type, spacecraft, orbits; resolutions; onboard vs. on-the-ground computing; inter-sensor communications, etc.
   - System being **designed in advance** as a mission or observing system **or incrementally and dynamically over time**
2. **Respond to various science and applied science events of interest:** Various overall observation timeframes; Various area coverages; Dynamic/Timely; Scheduling, re-targeting/re-pointing assets, as possible

System-of-Systems NOS-Testbed for technologies & concepts validation, demonstration, comparison and socialization

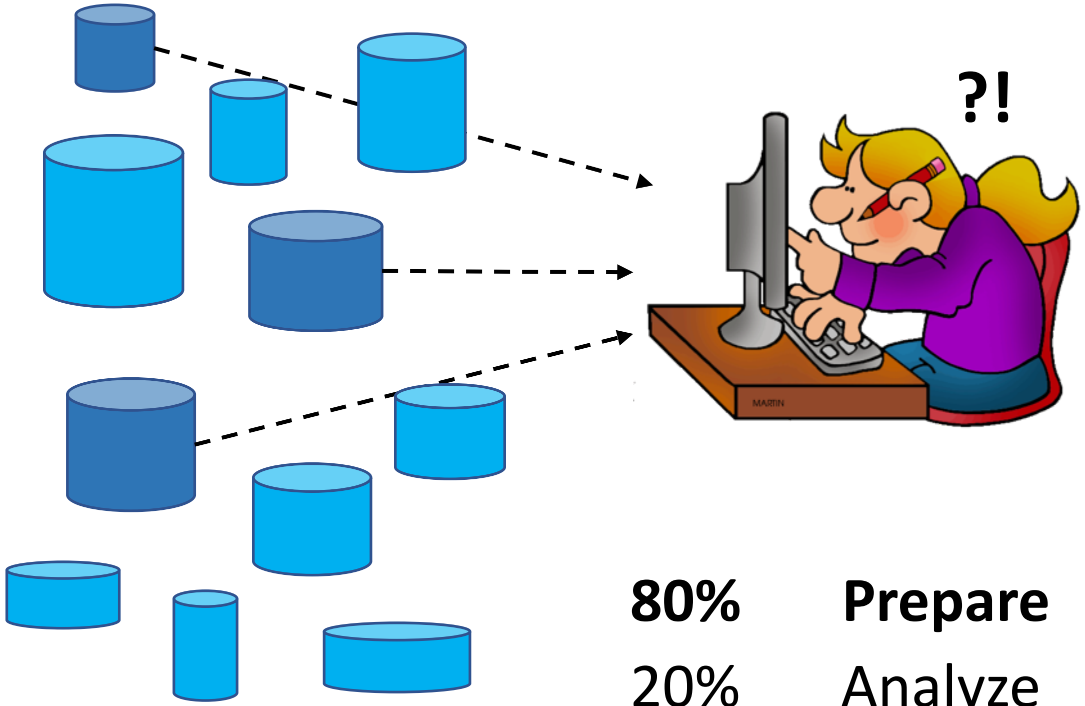
ESTO Earth Science Technology Office

From Archives to Analytic Centers:
Focus on the Science User

Data Archives

Focus on data capture, storage, and management

Each user has to find, download, integrate, and analyze

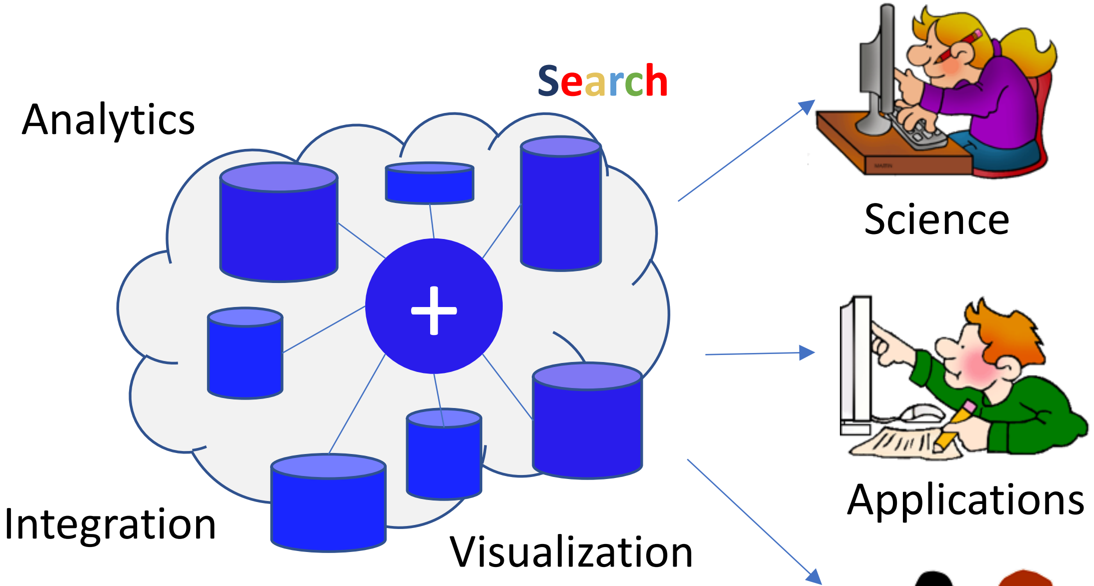


80%	**Prepare**
20%	Analyze

Analytic Centers

Focus on the science user

Integrated data analytics & tools tailored for a science discipline



20%	Prepare
80%	**Analyze**

Facilitates collaborative science across multiple missions and data sets

Analytic Collaborative Frameworks (ACF)

Focus is on the Science User

Allow flexibility/tailor configurations for Science investigators to choose among a large variety of datasets & tools

User

- Project Definition
- Plan for Investigation

Data

- Catalog
- NASA DAAC
- Other US Govt
- Non-US
- Local or non-public

Project Work Environment

Tools

- Discovery & Catalog
- Work Management
- Data Interfaces
- Analytic Tools
- Modeling
- Collaboration
- Visualization
- Sharing/ Publication
- Local/custom

Reduce repetitive work in data access and pre-processing, e.g., develop reusable components

Computational Infrastructure

- Computing
  - Capacity
  - Capability
- Storage
- Communications

Storage

- Data Containers
- Thematic model
- Metadata/Ontology
- Resulting Products
- Published data
- Provenance

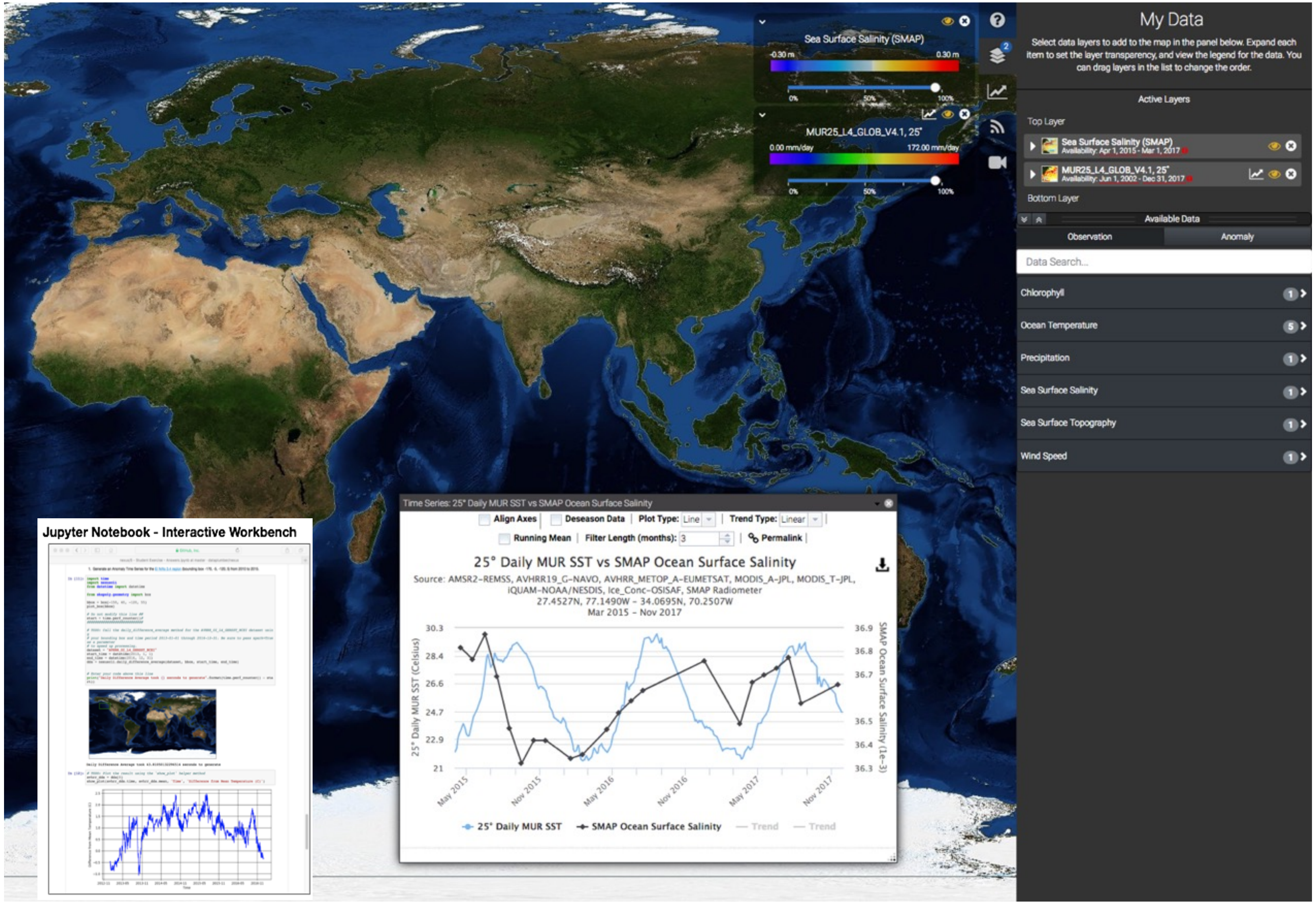
Computing

- Local systems
- High End Computing
- Cloud Computing Capability
- Quantum Computing
- Neuromorphic Computing

ESTO

Earth Science Technology Office

AIST OceanWorks:
An Analytic Collaborative Framework (ACF) for Ocean Science



Dozens of Ocean Data Sets
Stored in the cloud
Integrated, ready to use
Organized for fast search, subset and analysis

Search
Find relevant data sets

Cloud-based analytics
Analyze years of data over multiple data sets in seconds… without downloading data
Run analytics across multiple data sets despite differences in scale, times, granules, etc.

Custom analytics
Scientists can also run their own custom tools and algorithms

Integrated data
Match up in-situ and remote sensing data, despite differences in scale and resolution

Visualization
Subsets, layers, animations. Integrates with ArcGIS and Jupyter Notebooks

https://oceanworks.jpl.nasa.gov

Integration of NOS and ACF

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

Assimilate Observations

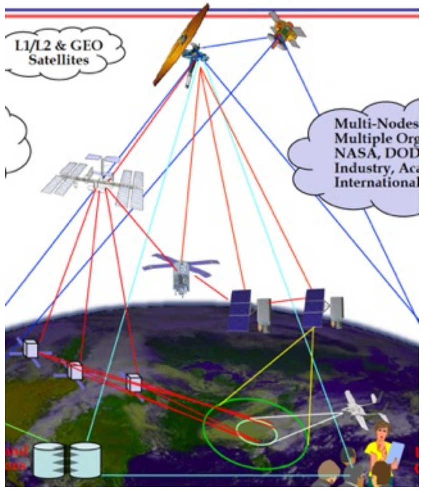
Enhance and enable focused Science investigations by facilitating access, integration and understanding of disparate datasets using pioneering visualization and analytics tools as well as relevant computing environments

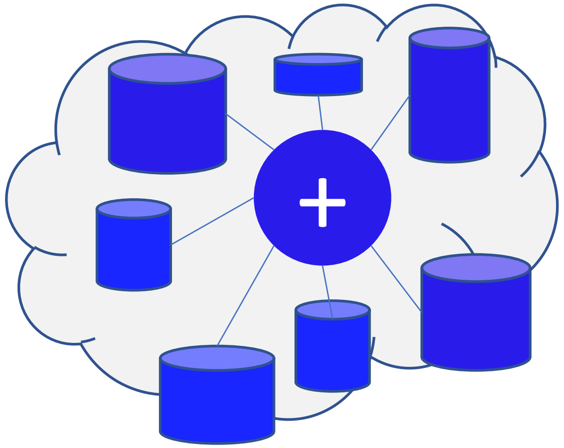
New Observing Strategies (NOS)

Analytic Collaborative Frameworks (ACF)

Acquire coordinated observations

Track dynamic and spatially distributed phenomena



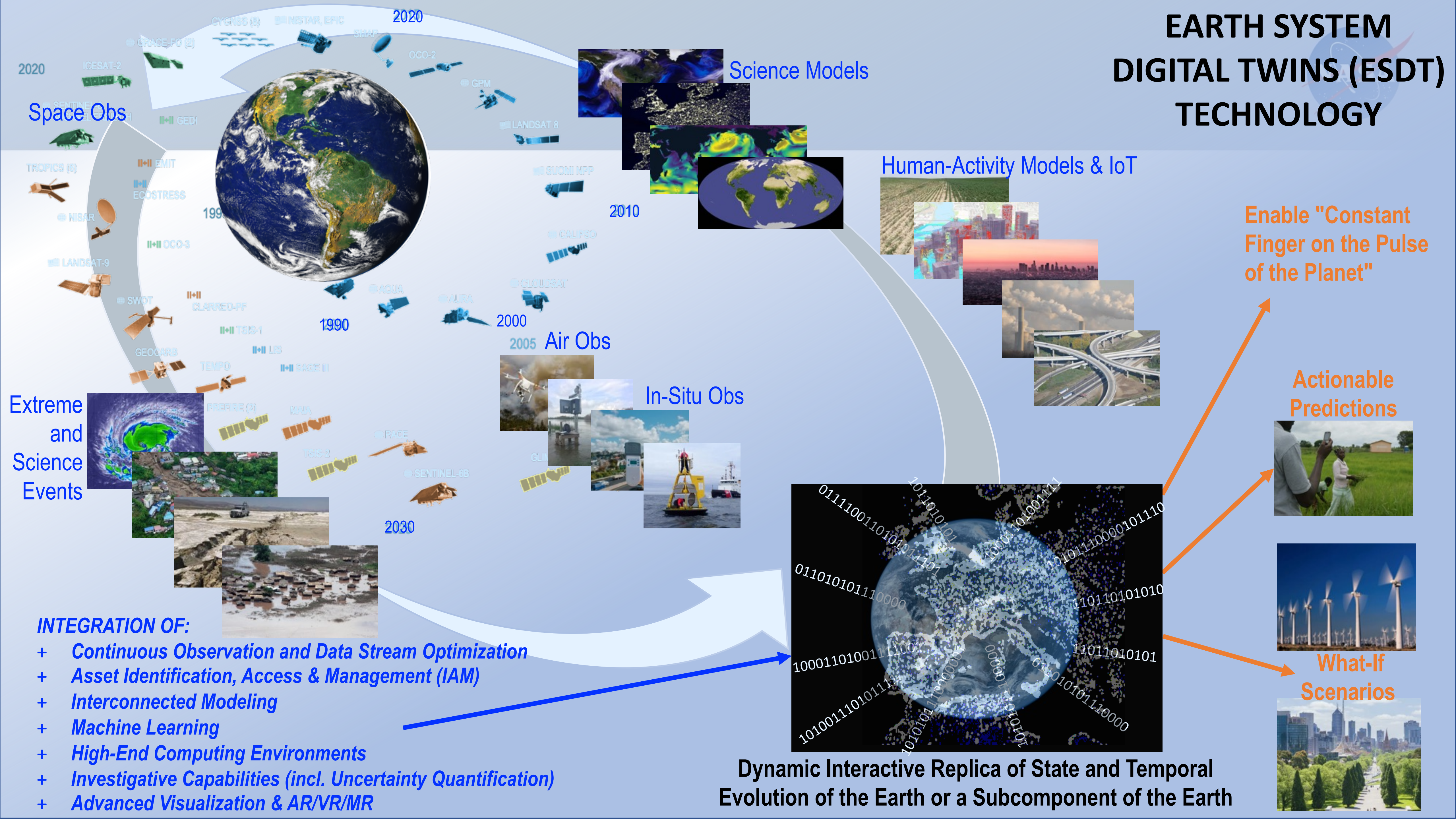


Assimilate many various data into models and analytic workflows.

What additional observations are needed?

Observation Requests

NOS+ACF acquires and integrates complementary and coincident data dynamically to build a more complete and in-depth picture of science phenomena


EARTH SYSTEM DIGITAL TWINS (ESDT) TECHNOLOGY
Space Obs
Science Models
Human-Activity Models & IoT
Air Obs
In-Situ Obs
Extreme and Science Events
Enable "Constant Finger on the Pulse of the Planet"
Actionable Predictions
What-If Scenarios
2020
2020
2010
2000
2005
1990
2030
ICESAT-2
OCO-2
GPM
LANDSAT 8
GEDI
EMIT
ECOSTRESS
NISAR
OCO-3
LANDSAT-9
SWOT
TSIS-1
TEMPO
SENTINEL-6B
INTEGRATION OF:
+ Continuous Observation and Data Stream Optimization
+ Asset Identification, Access & Management (IAM)
+ Interconnected Modeling
+ Machine Learning
+ High-End Computing Environments
+ Investigative Capabilities (incl. Uncertainty Quantification)
+ Advanced Visualization & AR/VR/MR
Dynamic Interactive Replica of State and Temporal Evolution of the Earth or a Subcomponent of the Earth

AI for Earth Science Applications

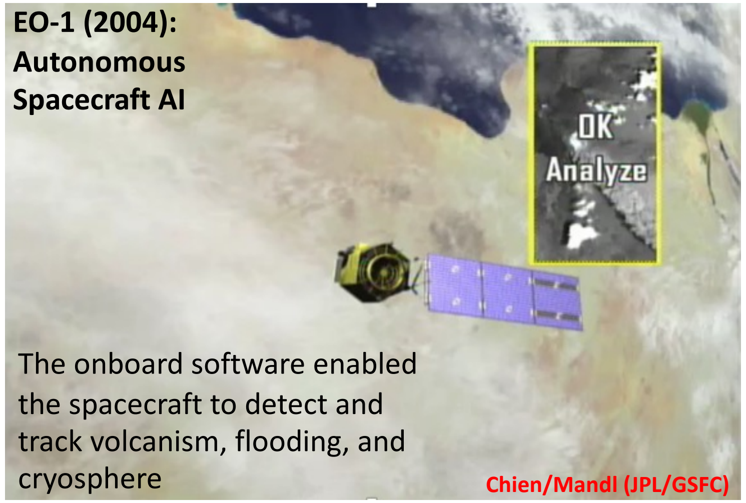
Two Main Areas

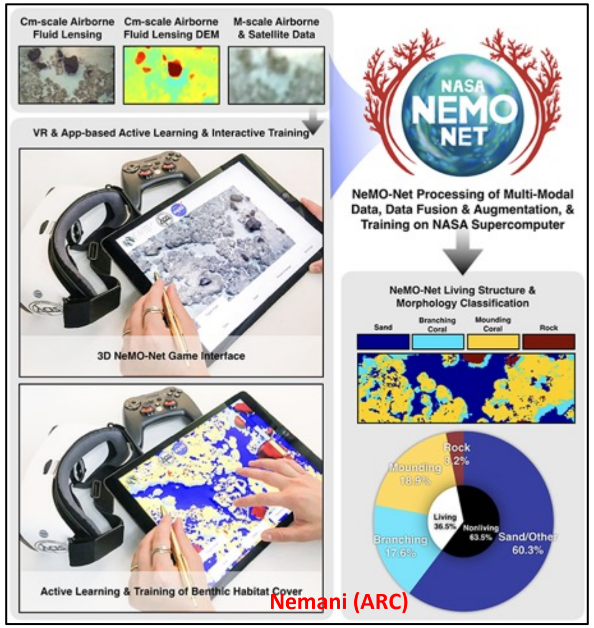
- **Improved Agile Observation Coordination and Mission Operations (Onboard or on the Ground)**
  - At the edge data analysis
  - Semi-autonomy and autonomy for decision making
  - Anomaly and fault detection
  - Engineering Support for large constellations
  - Advanced Interoperability

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, …





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)
- Trade-space Analysis Tool for Constellations Using Machine Learning (TAT-C ML) (Verville & Grogan/AIST-16)

AI for Time Series and for Science Models

- Advanced Phenology Information System (APIS) (Morisette/AIST-16)
- NASA Evolutionary Programming Analytic Center (NEPAC) (Moisan/AIST-18)
- Canopy Condition to Continental Scale Biodiversity Forecasts (Swenson/AIST-18)

AI for Quantum Computing

- Framework for Mining and Analysis of Petabyte-size Time-series on the NASA Earth Exchange (NEX) (Michaelis & Nemani/AIST-16)
- An Assessment of Hybrid Quantum Annealing Approaches for Inferring and Assimilating Satellite Surface Flux Data into Global Land Surface Models (Halem/AIST-16)

AI for Pattern and Information Extraction

- Computer-Aided Discovery and Algorithmic Synthesis for Spatio-Temporal Phenomena in InSAR (Pankratius/AIST-16)
- Autonomous Moisture Continuum Sensing Network (Entekhabi & Moghaddam/AIST-16)
- Supporting Shellfish Aquaculture in the Chesapeake Bay using AI for Water Quality (Schollaert-Uz/AIST-18)
- Mining Chained Modules in Analytics Center Frameworks (Zhang/AIST-18)

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

SpaceBorne-2 Experiments/Chien (NASA JPL) –
ISS Onboard Processing Experiment

This project will conduct both technology validation and TRL improvement experiments as well as it will demonstrate enhanced and enabling capabilities.

New Technology Demonstrations: *Validating New Hardware and Software Technologies*

- **Re-Tasking Demonstration :** using onboard data analysis to create alerts
  - Use onboard data analysis to generate alerts, NOS- or SensorWeb-like, e.g,: task other assets
- **Live-Instrument Data Feed :** run experiments with data generated on ISS, e.g., ECOSTRESS, EMIT, OCO-3
  - Using both pre-uploaded and potentially live data from onboard ISS instruments
- **Co-Processors Experiments:**
  - **Intel Movidius/Myriad Neuromorphic Processor**
    - Currently on ESA's Phi-Sat and terrestrial drones
  - **Qualcomm Snapdragon Processor**
    - Flying on Mars Helicopter
  - Gain tremendous in-space processing experience with 2 processors that are well on the path to mission use

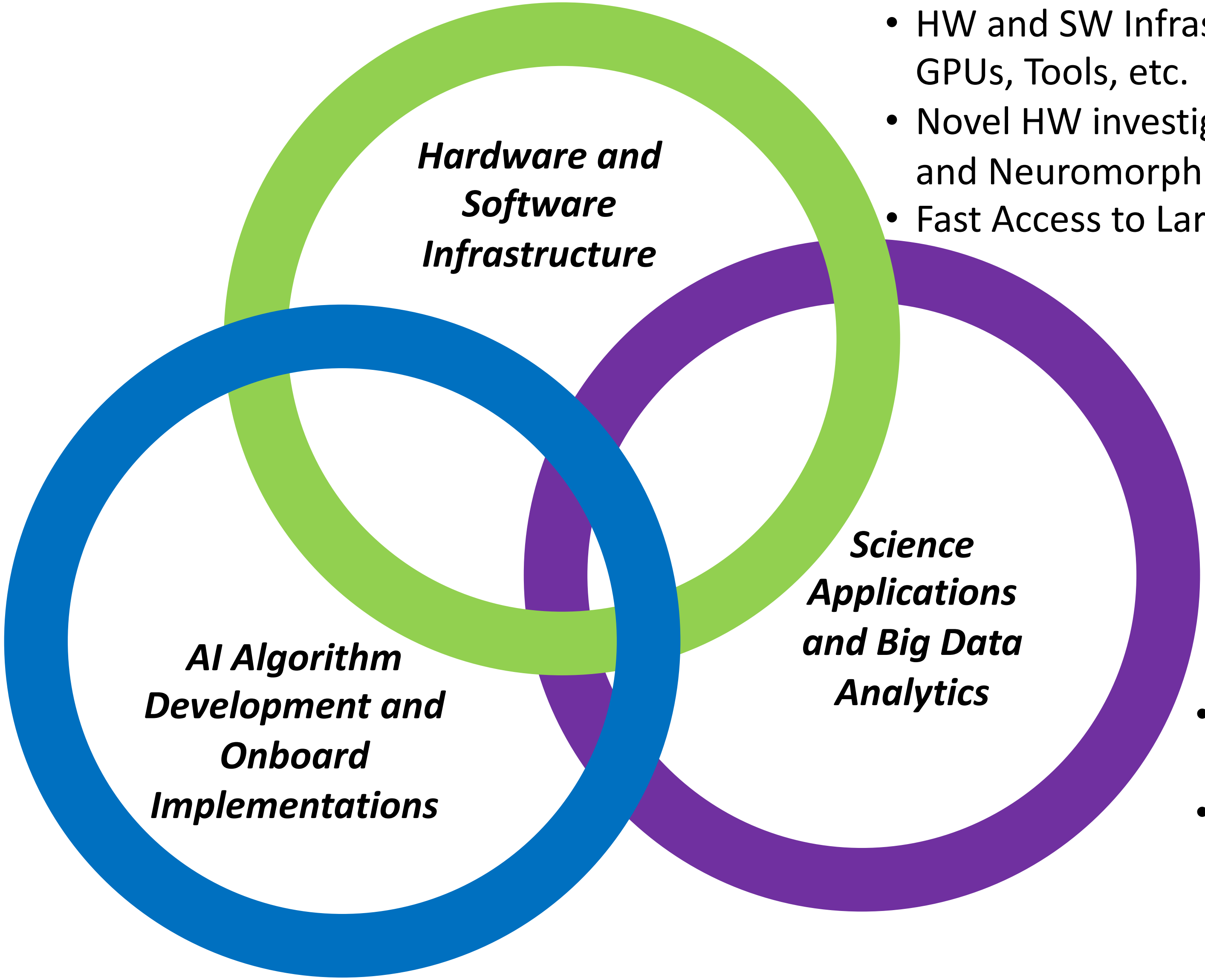
Data Processing and Machine Learning Experiments

- **Radar Processing: l**everage NISAR and UAVSAR radar pipelines => data reduction, low-latency downlink
- **Thermal Infrared Processing :** experiment with TIR data from ECOSTRESS
  - Onboard pipeline: radiometric calibration, geolocation, land surface temperature, etc.
  - Applied Science Value: orders of magnitude data reduction, low-latency downlink
- **VSWIR Processing :** experiment with VSWIR data from EMIT
  - Heritage technology from HysPIRI Intelligent Payload Module (IPM)
  - Applied Science Value: data reduction, low-latency downlink, alerts for tasking other sensors
- **Machine Learning Demonstration** : perform ML/imagery classification techniques like HiRISNet, MSLnet, and Hirise

AI for Space Applications
Current & Future

- 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


Any Questions?

AIST-18/Beck (U. Alabama, Huntsville) –

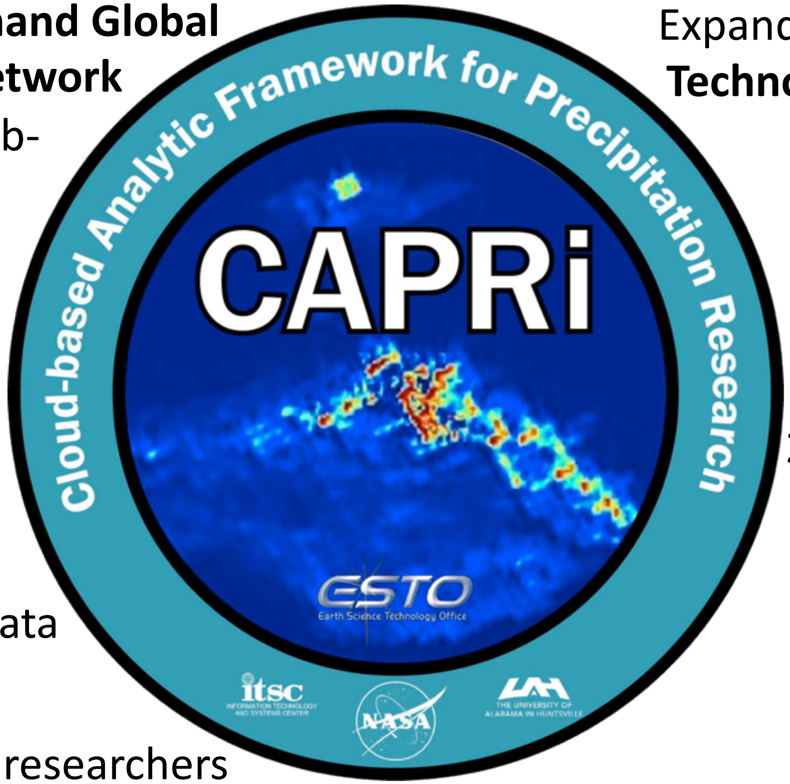
Cloud-Based Analytic Framework for Precipitation Research

Leverage cloud-native technologies from the AIST-2016 VISAGE project to develop a Cloud-based ACF for Precipitation Research using a Deep Learning (CNNs) framework to provide an analysis-optimized cloud data store and access via on-demand cloud-based serverless tools . It uses coincident ground and space radar observations.

Provide users with tools for **on-demand Global Precipitation Mission Validation Network (GPM VN) data** querying, fusion, sub-setting, extraction, and analysis integrated with Deep Learning architectures.

Develop **Super-Resolution** of remotely sensed images and develop **higher-resolution product** based on GPM Dual-frequency Precipitation Radar (DPR) gridded data prototype.

Ability to generate training data for researchers to use with Deep Learning models



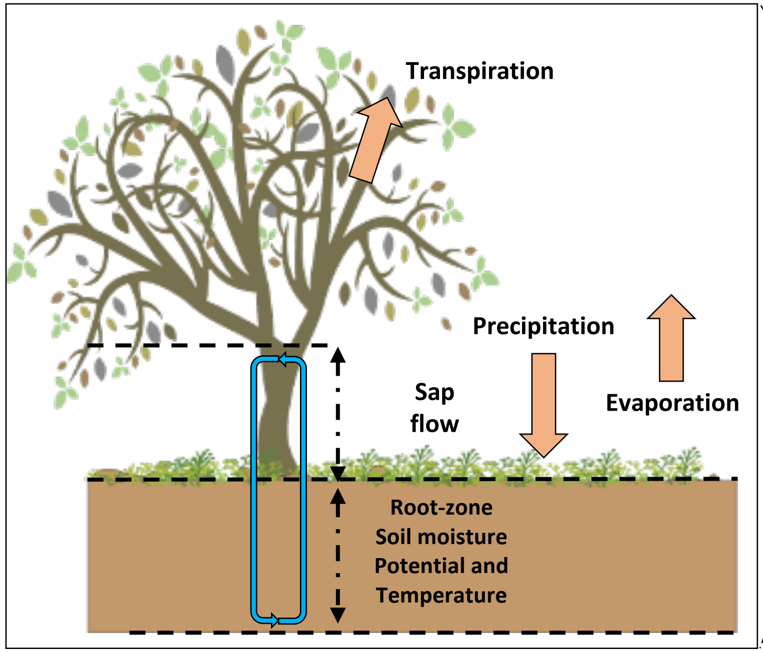
Expand state of knowledge in **Cloud Technologies for precipitation data** by:

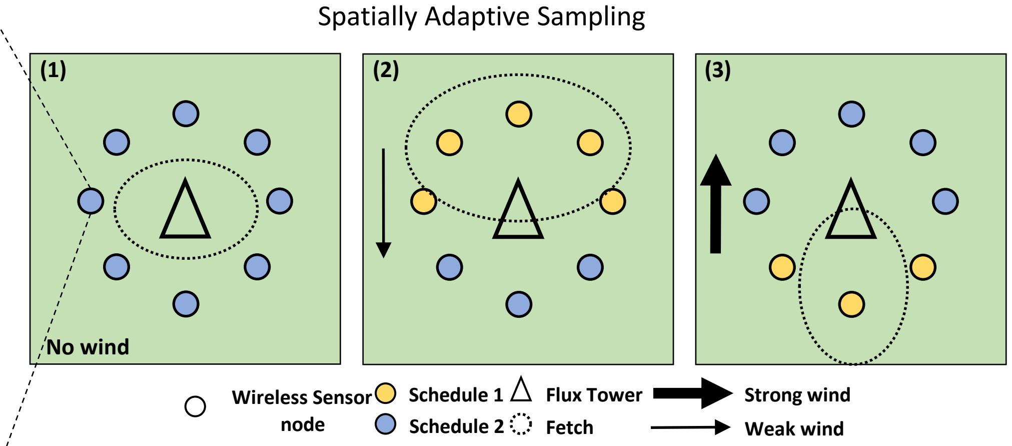
1. Providing new solutions for **real-time querying** of large data sets in a serverless environment;
2. Developing new methods for **generating Deep Learning Training data** on the fly; and
3. Providing an **easy-to-use user interface** for analysis and visualization of the data.

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

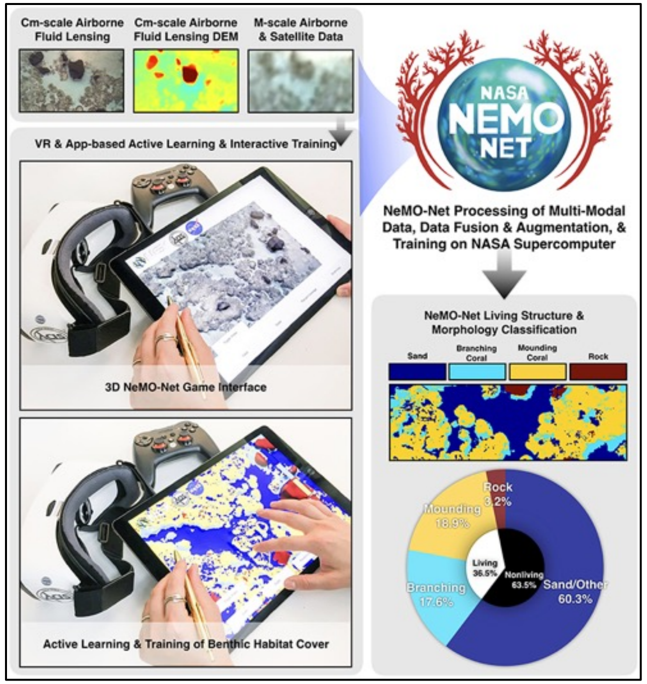
SoilSCAPE installation for CYGNSS Cal/Val

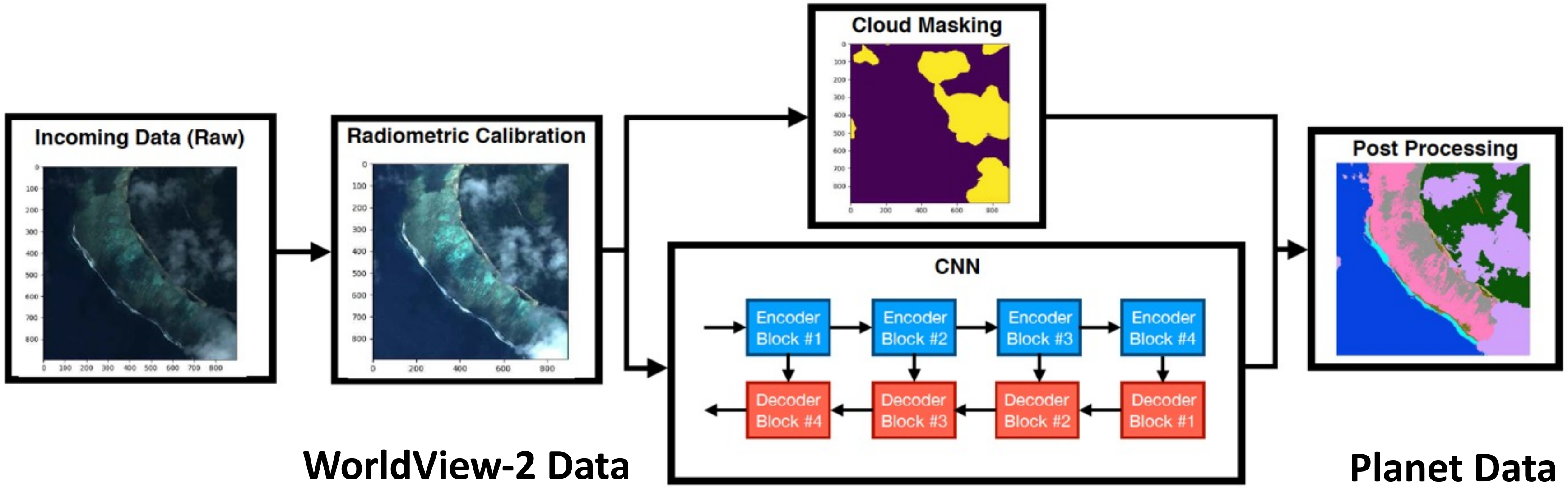
- SoilSCAPE Plan → Satellites Cal/Val
  - SMAP Cal/Val: Deployed 1 site at the Cary Institute of Ecosystem Studies (Millbrook, NY)
  - SoilSCAPE team (via. Co-I Moghaddam) collaborating with CYGNSS to provide *in situ* soil moisture for cal/val activities
  - Established a cal/val infrastructure for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

AIST-16/Chirayath (ARC) –

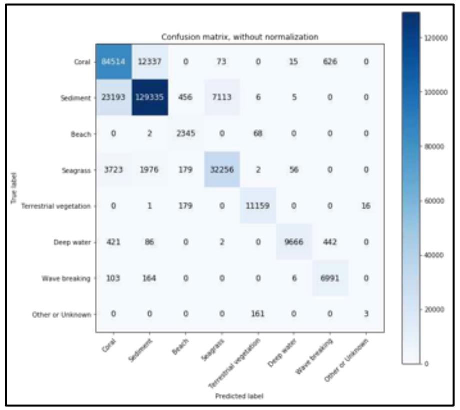
NeMO-Net - The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment

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.



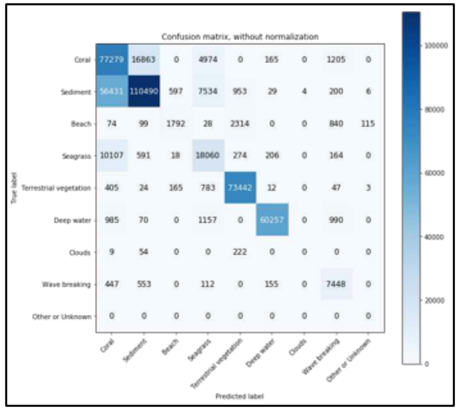


WorldView-2 Data



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

Planet Data



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

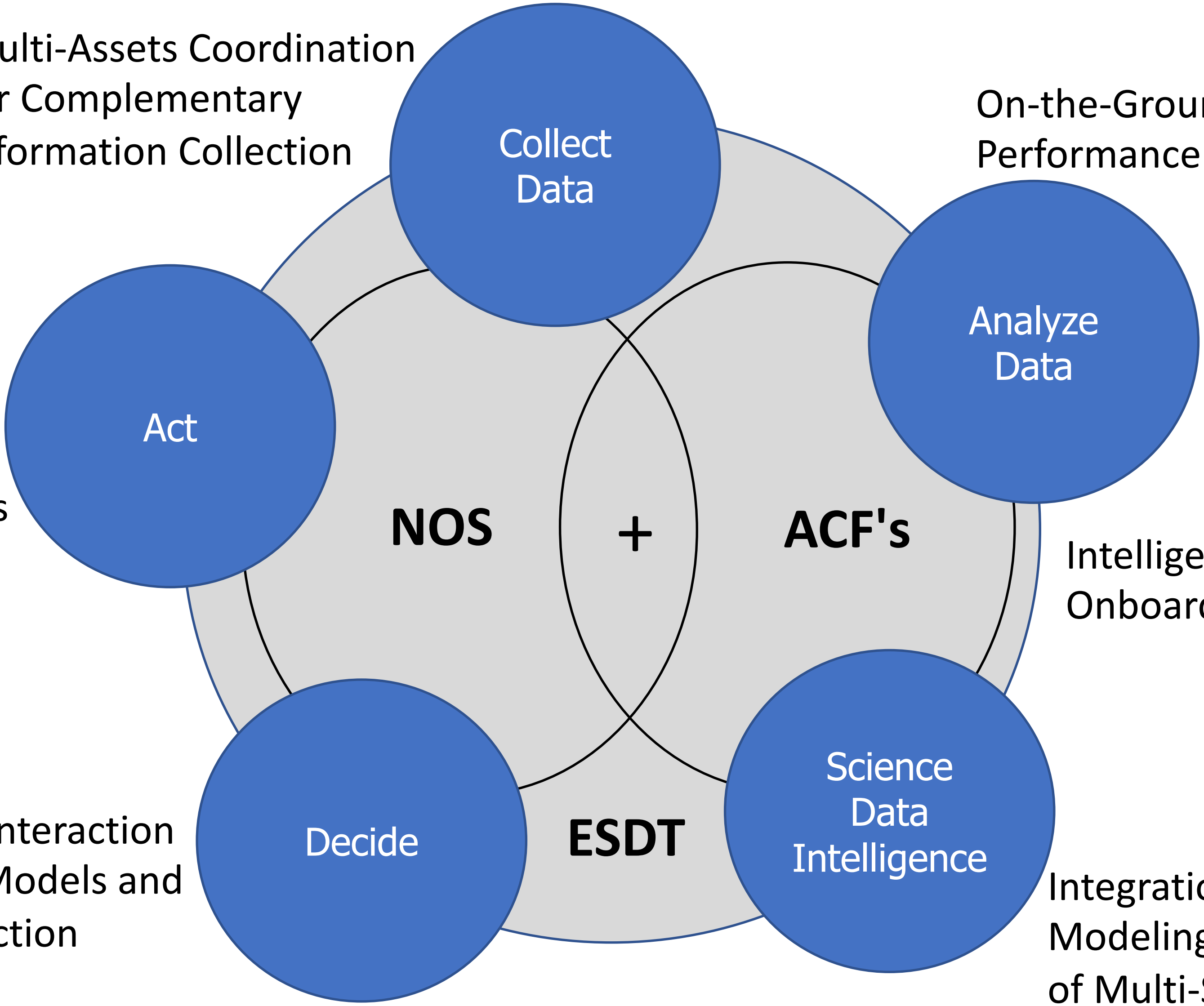
Integrated Earth System Science

Continuous Integration of NOS and ACF Capabilities

Multi-Assets Coordination for Complementary Information Collection

On-the-Ground and Onboard High-Performance Computing Capabilities

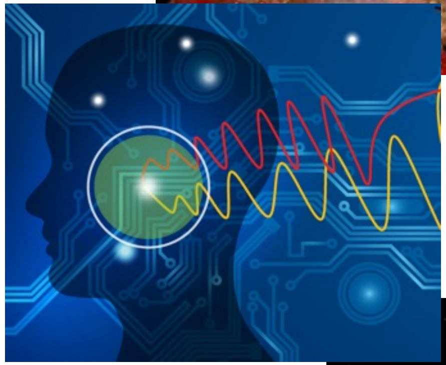
Dynamic Reaction to Science Events of Interest with Varying Latency Requirements

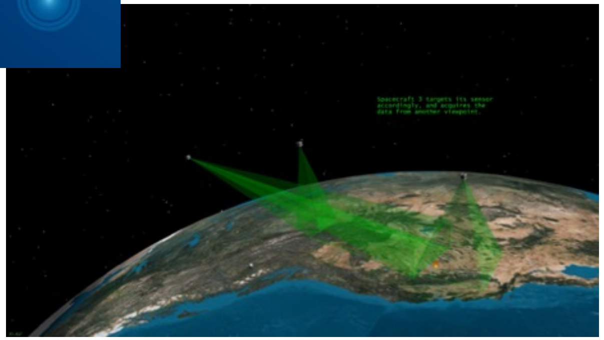


Intelligent Data Analytics Onboard and On-the-ground

Seamless Interaction between Models and Data Collection

Integration, Fusion, Visualization, Modeling and Analysis of Large Amounts of Multi-Source Data





Back-up Slides (Projects Details)

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.

NOS Application Cases

Mission Type ***Timeframe*** ***Application***	**Tactical Observing System** ***Seconds-minutes*** ***Point event/phenomenon***	**Operational Observing System** ***Hours-days*** ***Spatial phenomenon***	**Strategic Observing System** ***Months-years*** ***Spatial-temporal phenomenon***
Example	*Detect and observe volcanic activity*	*Increase spatial observation of primary forest burning as input into long-term Air Quality and Climate models*	*Select observing strategy to optimize all measurements that will improve hydrologic estimates*
Functions	Detect emergent event Deploy observation assets	Deploy observation assets Digest information sources	Design observation system Digest information sources
Capabilities	• Responsiveness • Interaction • Dynamics • Adaptation	• Resource allocation • Coordination • Data assimilation • Prediction/ forecasting	• Platform selection • Coordination • Data assimilation • State estimation (belief)

AIST Awards – NOS Clusters

- **NOS-T Relevant**

PI's Name	Organization	Title	Synopsis
Mahta Moghaddam	U. of Southern California	SPCTOR: Sensing Policy Controller and OptimizeR	Multi-sensor coordinated operations and integration for soil moisture, using ground-based and UAVs "Sensing Agents".
Jim Carr	Carr Astro	StereoBit: Advanced Onboard Science Data Processing to Enable Future Earth Science	SmallSat/CubeSat high-level onboard science data processing demonstrated for multi-angle imagers, using SpaceCube processor and CMIS Instrument, and Structure from Motion (SfM).
Sreeja Nag	NASA ARC	D-SHIELD: Distributed Spacecraft with Heuristic Intelligence to Enable Logistical Decisions	Suite of scalable software tools - Scheduler, Science Simulator, Analyzer to schedule the payload ops of a large constellation based on DSM constraints (mech, orb), resources, and subsystems. Can run on ground or onboard.
Paul Grogan	Stevens Institute of Technology	Integrating TAT-C, STARS, and VCE for New Observing Strategy Mission Design	Inform selection and maturation of Pre-Phase A distributed space mission concept, by integrating: TAT-C: architecture enumeration and high-level evaluation (cost, coverage, quality); STARS: autonomous/adaptive sensor interaction (COLLABORATE); VCE: onboard computing and networking

- **OSSEs (Observing System Simulation Experiments)**

PI's Name	Organization	Title	Synopsis
Derek Posselt	NASA JPL	Parallel OSSE Toolkit	Fast-turnaround, scalable OSSE Toolkit to support both rapid and thorough exploration of the trade space of possible instrument configurations, with full assessment of the science fidelity, using cluster computing.
Bart Forman	U. of Maryland	Next Generation of Land Surface Remote Sensing	Create a terrestrial hydrology OSSE/mission planning tool with relevance to terrestrial snow, soil moisture, and vegetation using passive/active microwave RS, LiDAR, passive optical RS, hydrologic modeling, and data assimilation, using LIS and TAT-C.
Ethan Gutmann	UCAR	Future Snow Missions: Integrating SnowModel in LIS	Improve NASA modeling capabilities for snow OSSE, to plan and operate a future cost-effective snow mission by coupling the SnowModel modeling system into NASA LIS.

Earth Science Technology Office

NOS-T Architecture and Pilot Projects

PI's Name	Organization	Emails	Title	Synopsis
Tom McDermott & Paul Grogan & Jerry Sellers	Systems Engineering Research Center (SERC)	tmcdermo@stevens.edu; pgrogan@stevens.edu; jsellers@tsti.net	New Observing Strategies Testbed (NOS-T) Design and Development	Design the NOS-T framework to enable system-of systems experiments and testing; enable multi-party and geographically distributed participation and connected tests and operations; enables both open community and protected exchange of measurement data; provide a communications infrastructure; and simulate actual operational security challenges.
Chad Frost & Daniel Cellucci	NASA Ames	chad@nasa.gov; daniel.w.cellucci@nasa.gov	Earth Science "Tip and Cue" Technologies for a New Observing Strategy	Extend the capabilities of the Generalized Nanosatellite Avionics Testbed (G-NAT) and networked, state-of-the-art, miniaturized, tracking and sensing devices (termed 'tags'), developed in collaboration with USGS, to enable a tip-and-cue architecture for dynamically reconfigurable remote sensing.
Sujay Kumar & Rhae Sung Kim	NASA Goddard	sujay.v.kumar@nasa.gov; rhaesung.kim@nasa.gov	A Hydrology MIssion Design and Analysis System (H-MIDAS)	Extend LIS capabilities to: support the incorporation of distributed sensor observations fir hydrology; support the development of observation operators; perform data assimilation simulations and provide feedback to the observing systems.
Steve Chien & James Mason	NASA JPL	steve.a.chien@jpl.nasa.gov; james.mason@jpl.nasa.gov	Planning and Scheduling for Coordinated Observations	Develop a planning and scheduling framework for the NOS Testbed that will coordinate multiple observing assets (e.g. space, air, land) to perform coordinated and continuous measurements at varying scales (e.g. spatial, temporal).
Dan Crichton & Cedric David	NASA JPL	daniel.j.crichton@jpl.nasa.gov; cedric.david@jpl.nasa.gov	NOS Testbed Study and Science Use Cases Identification	Contribute to the definition of the NOS Testbed by identifying science use cases, observing assets, requirements, interfaces, and other design recommendations in close collaboration with the NOS Testbed Definition activity.
Louis Nguyen	NASA LaRC		Ground Stations as a Service (GSaS) for Near Real-time Direct Broadcast Earth Science Satellite Data	Utilize GSaS to receive direct broadcast (DB) data from EOS to significantly reduce latency in acquiring LEO satellite observations (e.g., from 3-6 hours to 20-25 mins). It will provide ability to receive low latency LEO data without the need to own/maintain DB ground station; improve NASA Earth Science's ability to deliver lower latency products and therefore increasing optimal use; provide NOS with capability to schedule, coordinate, receive, and process DB data from EOS.
Jay Ellis	KBR/GSFC	nathaniel.j.ellis@nasa.gov	NOS Testbed Administration and Management	Administer and manage the NOS Testbed for disparate organizations to propose and participate in developing NOS software and information systems technology capabilities and services.

AIST18 Awards – ACF Clusters

- **Biodiversity ACF**

Award #	PI's Name	Organization	Title	Synopsis
AIST-18-0007	Schollaert Uz	NASA GSFC	Supporting shellfish aquaculture in the Chesapeake bay using AI for water quality	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.
AIST-18-0031	Moisan	NASA GSFC	NASA Evolutionary Programming Analytic Center (NEPAC)	Discover and apply novel algorithms for ocean chlorophyll using AI/ML (Genetic Programming) on satellite/in-situ obs and a user-friendly GUI to connect data and applications with HEC resources for improved science.
AIST-18-0034	Jetz	Yale U.	Biodiversity - Environment Analytic Center	Near real-time monitoring of the biological pulse of our planet, using an online dashboard, taking into account various spatiotemporal resolutions, data uncertainty and biodiversity data biases, and supporting analysis, visualization and change detection across scales.
AIST-18-0043	Townsend	U. Wisconsin, Madison	GeoSPEC: On-Demand Geospatial Spectroscopy Processing Environment on the Cloud	Develop a framework/processing workflow for on-demand cloud-based Hyperspectral/Spectroscopy Science Data Processing in preparation for SBG needs. Will provide options for new atmospheric & other types of corrections, possibilities for users' or commercial code. Will be tested with AVIRIS-Classic and –NG data.
AIST-18-0063	Swenson	Duke University	Canopy condition to continental scale biodiversity forecasts	Characterize canopy condition from various spatio-temporal RS products (including drought indices and habitat structure) to predict supply of mast resources to herbivores (and threatened species) and visualize canopy condition and drought-stress maps

- **Land Cover ACF**

Award #	PI's Name	Organization	Title	Synopsis
AIST-18-0020	Ives	U. Of WI, Madison	Valid time series analyses for satellite data	Develop new statistical tools to analyze large, time series of various remotely sensed datasets and provide statistical rigor and confidence to conclusions about patterns of change and to forecasts of future change, identifying patterns of annual trends, seasonal trends and phenological events, and analyzing the cause of these trends.

AIST18 Awards – ACF Clusters (cont.)

- **Air Quality ACF**

Award #	PI's Name	Organization	Title	Synopsis
AIST-18-0011	Martin	Washington U.	Development of GCHP to enable broad community access to high-resolution atmospheric composition modeling	Integrate atmospheric chemistry models online into Earth system models (ESMs) and offline using meteorological data, using the high-performance version of the GEOS-Chem global 3-D model of atmospheric chemistry (GCHP) and the Earth System Modeling Framework (ESMF) in its Modeling Analysis and Prediction Layer (MAPL) implementation.
AIST-18-0044	Duren	NASA JPL	Multi-scale Methane Analytic Framework	ACF for methane data analysis spanning multiple observing systems and spatial scales with workflow optimization, analytic tools to characterize methane fluxes and physical processes, tools for data search and discovery, and a collaborative, web-based portal.
AIST-18-0072	Henze	U. of CO, Boulder	Surrogate modeling for atmospheric chemistry and data assimilation	Advance computational tools available for AQ prediction, mitigation, and research by building a robust and computationally efficient chemical Data Assimilation system, merging research in compressive sampling and machine learning for large-scale dynamical systems and integrating multi-source data into an existing model.
AIST-18-0099	Holm	City of Los Angeles	Predicting What We Breathe: Using Machine Learning to Understand Urban Air Quality	Link ground-based in situ and space-based remote sensing observations of major AQ components to classify patterns in urban air quality, enable the forecast of air pollution events, and identify similarities in AQ regimes between megacities around the globe, using science models and ML-based algorithms.

- **Precipitation ACF**

Award #	PI's Name	Organization	Title	Synopsis
AIST-18-0051	Beck	U. Of AL, Hunstville	Cloud-based Analytic Framework for Precipitation Research	Leverage cloud-native technologies from the AIST-2016 VISAGE project to develop a Cloud-based ACF for Precipitation Research using a Deep Learning (CNNs) framework to provide an analysis-optimized cloud data store and access via on-demand cloud-based serverless tools . It will use coincident ground and space radar observations.

AIST18 Awards – ACF Clusters (cont.)

• Disaster Management ACF

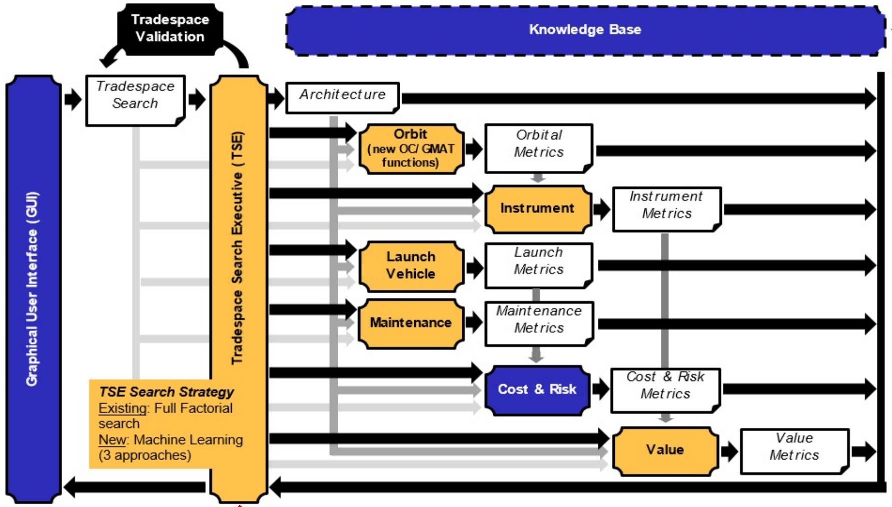
Award #	PI's Name	Organization	Title	Synopsis
AIST-18-0055	Coen	NCAR	Creation of a Wildfire Fire Analysis: Products to Enable Earth Science	Develop methods to create, test and assess wildland fire reanalysis products (standardized, gridded wildland fire information generated at regular intervals) using fire detection data, as well as coupled weather-wildland fire model and data assimilation.
AIST-18-0001	Donnellan	NASA JPL	Quantifying Uncertainty and Kinematics of Earthquake Systems ACF (QUAKES-A)	Create a uniform crustal deformation reference model for the active plate margin of California by fusing data with widely varying spatial and temporal resolutions, quantifying uncertainty, developing data management and geospatial information services and providing collaboration and infusion into target communities.
AIST-18-0085	Hua	NASA JPL	Smart On-Demand of SAR ARDs in Multi-Cloud & HPC	Enable full resolution time series analysis, high-accuracy flood and damage assessments with remote sensing SAR Analysis Ready Data (ARD), using Jupyter Notebooks and on-demand analysis across multi-cloud environments.

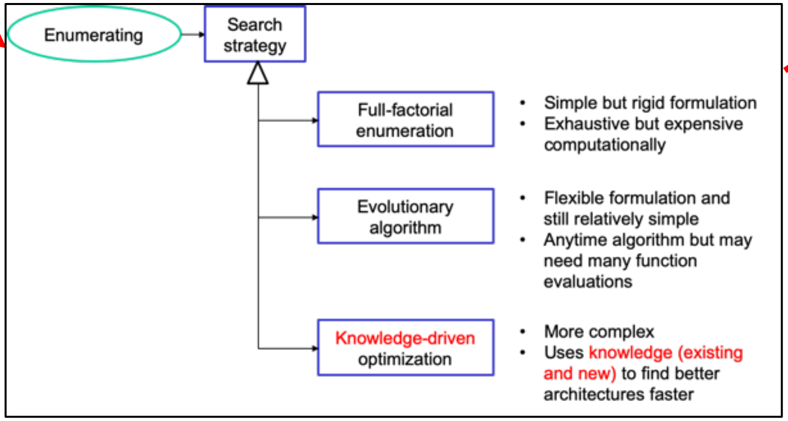
• Cross-Cutting ACF Capabilities

Award #	PI's Name	Organization	Title	Synopsis
AIST-18-0042	Huffer	Lingua Logica	AMP: An Automated Metadata Pipeline	Automate and improve the use and reuse of NASA Earth Science data by developing a fully-automated metadata pipleline integrating ML and ontologies (SWEET) for a semantic, metadata mining from data. Developed in collaboration with GES DISC.
AIST-18-0059	Zhang	Carnegie Mellon U.	Mining Chained Modules in Analytics Center Framework	Build a workflow tool as a building block for ACF, 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, develop algorithms to extract reusable chain of software modules, and develop an intelligent service that provides for personalized recommendations.

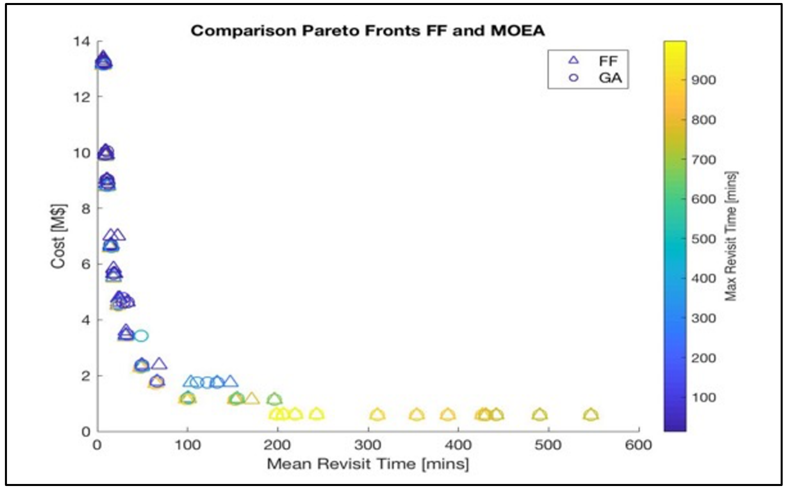
AIST-14 & -16/Grogan (Stevens) –
Trade-space Analysis Tool for Constellations (TAT-C)

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]

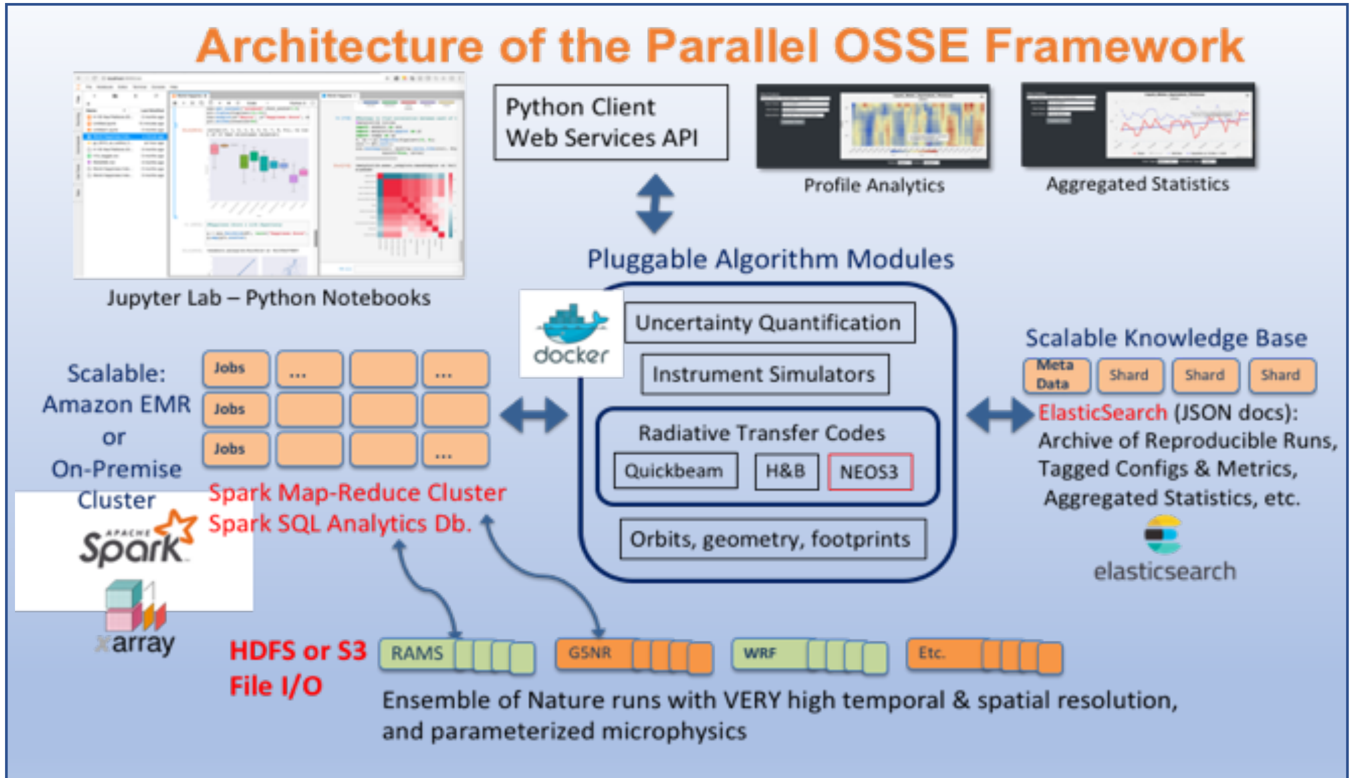
Note: Previous PI's: Le Moigne & Verville @GSFC

AIST-18/Posselt (NASA JPL) –
Parallel Observing System Simulation Experiment (OSSE) Toolkit

Fast-turnaround, scalable OSSE Toolkit to support both rapid and thorough exploration of the trade space of possible instrument configurations, with full assessment of the science fidelity, using cluster computing.

Evaluate measurement contribution to mission science

- **Parallel OSSE system** uses distributed computing, data analytics, and Bayesian retrieval algorithms to rapidly and thoroughly evaluate information in a wide variety of prospective measurements



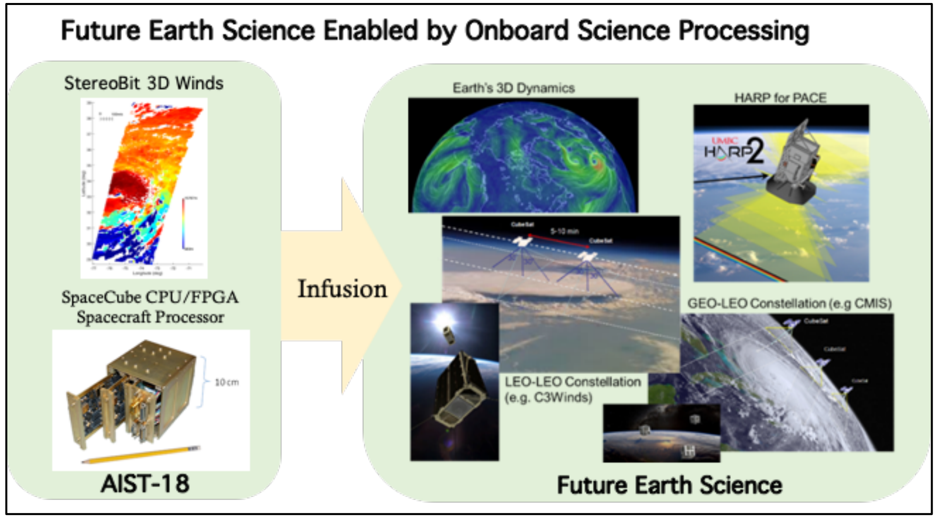
Includes: Interfaces to Nature Run databases; Pluggable instrument simulators and retrieval algorithms; Quantitative evaluation of science benefit vs SATM

Technology already proven for evaluation of measurement goals for the ACCP DO study

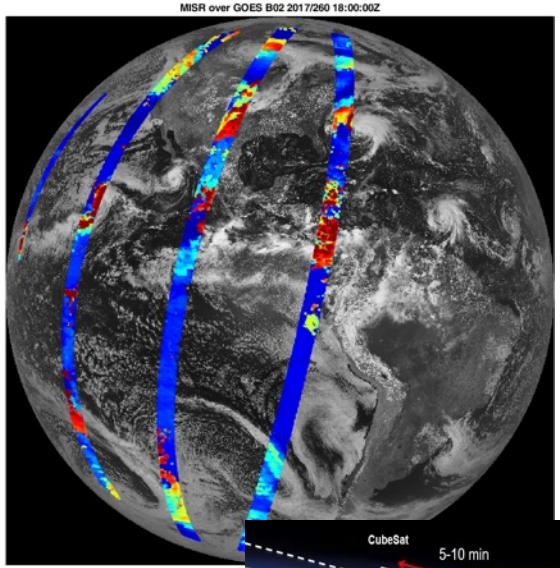
- Couple **instrument simulators** with a scalable parallel computing framework utilizing **Map-Reduce compute cluster**, a **scalable Knowledge Base**, a set of "pluggable" code modules, and **Python Live Notebooks**
- Produce **quantitative estimates of geophysical variable uncertainty** and information on mission architecture sufficiency
- Considered to address PBL observational goals

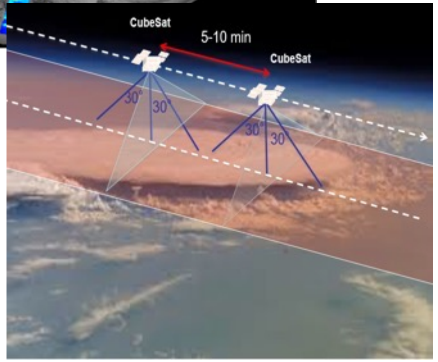
AIST-18/Carr (Carr Astronautics) –
StereoBit: Advanced Onboard Science Data Processing to Enable Future Earth Science

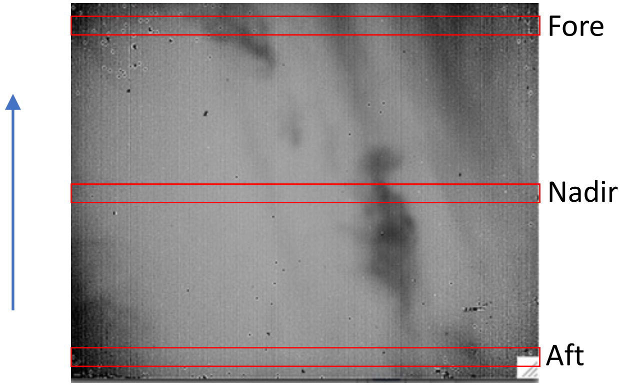
This investigation will demonstrate higher-level onboard science data processing for more intelligent SmallSats and CubeSats to enable future Earth science missions and Earth observing constellations. Low-cost SmallSat architectures generally suffer from downlink bottlenecks and often result in lower 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)

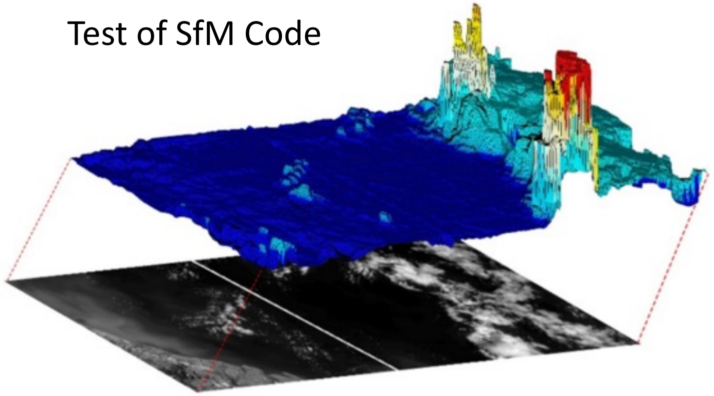






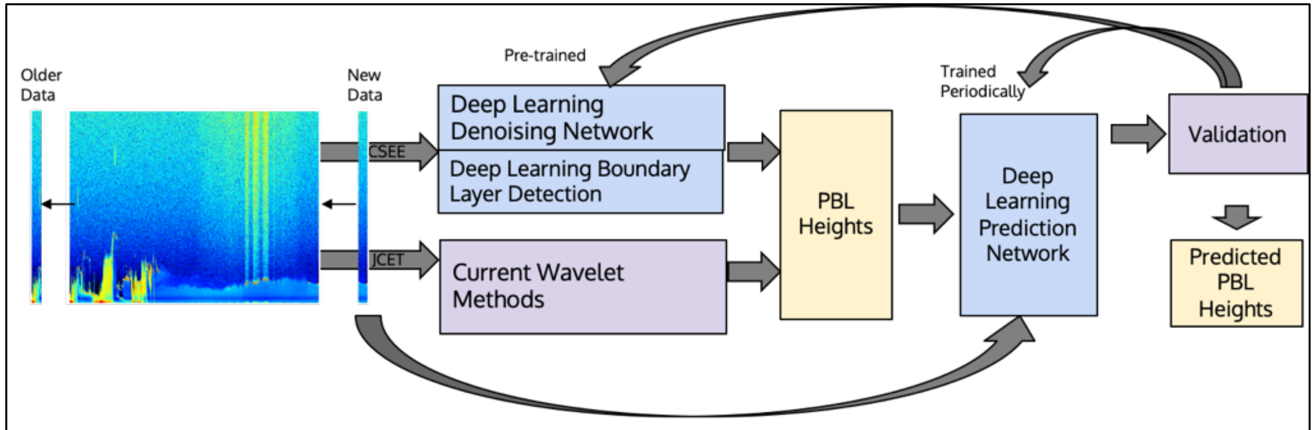
Early CTI Cloud Picture

Test of SfM Code

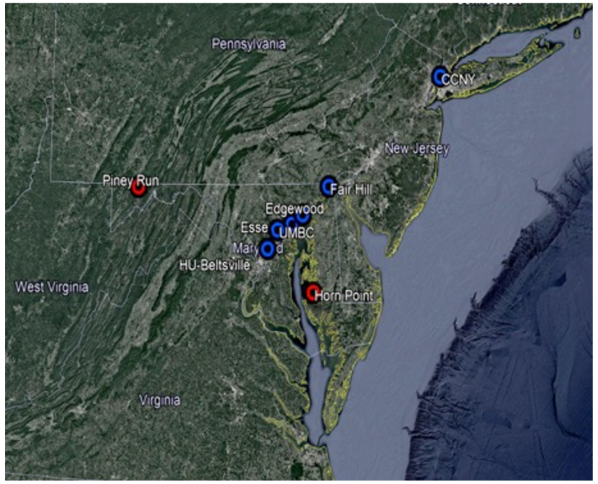


AIST-16/Halem (UMBC) – *Implement a Deep Learning Operational Ceilometer (LIDAR)-based Atmospheric Boundary Layer Height Product over Continental U.S.*

This project will prototype an operational Atmospheric Boundary Layer Heights (ABLH) network product based on deep learning segmentation algorithms. A deep learning edge detection algorithm trained on a small data set of completely unrelated ABLH images was implemented and that detected consistent ABLH with the double derivative method. When ML images were overlayed on backscatter plot with radiosonde validation points, it showed better accuracy than the double derivative method..



Integration of Machine Learning Methods for Operational PBLH



Current Ceilometer/Lidar Sites

- Fair Hill: Vaisala CL31
- Edgewood: Vaisala CL51
- Essex: Lufft CHM8k
- UMBC: Lufft CHM15k and MPL Sigma Space
- Beltsville (Howard Univ.): Lufft CHM15k
- CCNY: Lufft CHM15k

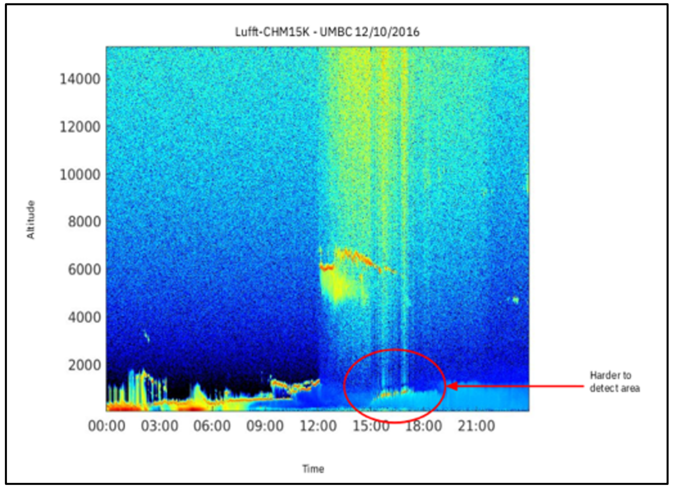
Radar Wind Profiler

- HU-Beltsville
- Horn Point Laboratory
- Piney Run

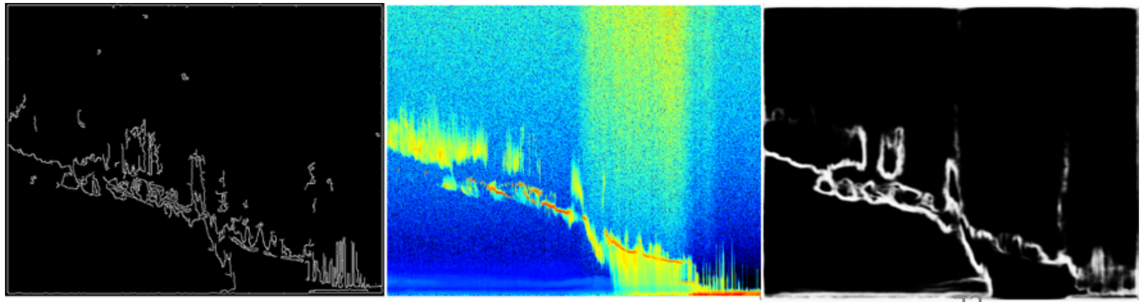
Current methods can be problematic under the following conditions:

- When residual layers are present
- Rain or cloudy conditions exist

Under these conditions, current methods can result in gaps in coverage or discrepancies in estimations when compared with radiosondes.



Comparing Deep Learning Boundary Detection with Standard Edge Detection Methods



Canny Edge Detection | LIDAR Backscatter over 24 hours (Lufft-CHM15K Dec 6, 2016) | Deep Neural Network Edge Detection

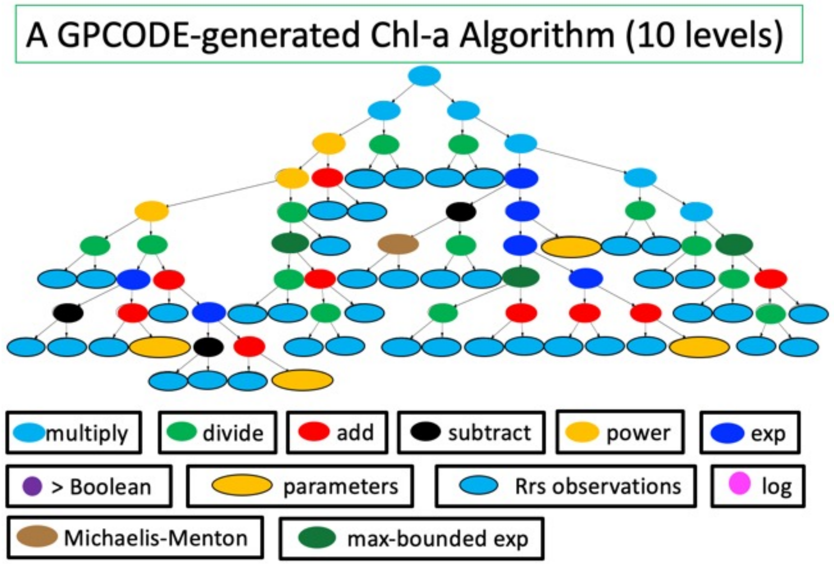
ESTO Earth Science Technology Office

AIST-18/Moisan (GSFC) and Zhang (SMU) –
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)

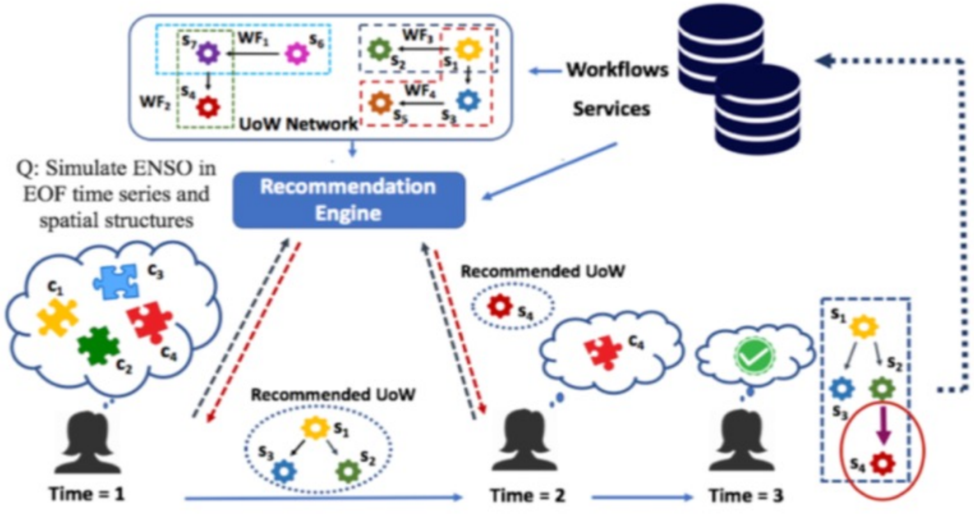


Improved Ocean Color Algorithms and Climate Data Records Using Machine Learning/Genetic Programming

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

AIST Earth System Digital Twin(s)

AIST defines an Earth System Digital Twin (ESDT) as an interactive and integrated multidomain, multiscale, digital replica of the state and temporal evolution of Earth systems that dynamically integrates:

- Relevant Earth system models and simulations
- Other relevant models (e.g., related to the world's infrastructure); continuous and timely (including near real time and direct readout) observations (e.g., space, air, ground, over/underwater, Internet of Things (IoT), socioeconomic)
- Long-time records
- Analytics and artificial intelligence tools.

Effective ESDTs enable users to run hypothetical scenarios to improve the understanding, prediction of and mitigation/response to Earth system processes, natural phenomena and human activities as well as their many interactions.
An ESDT is a type of integrated information system that, for example, enables continuous assessment of impact from naturally occurring and/or human activities on physical and natural environments.

AIST ESDT strategic goals are to:

- Develop information system frameworks to provide continuous and accurate representations of systems as they change over time;
- Mirror various Earth Science systems and utilize the combination of Data Analytics, Artificial Intelligence, Digital Thread*, and state-of-the-art models to help predict the Earth's response to various phenomena;
- Provide the tools to conduct "what if" investigations that can result in actionable predictions.

* The digital thread designates the communication framework that links all digital twin data flow throughout its lifecycle.

ESTO
Earth Science Technology Office