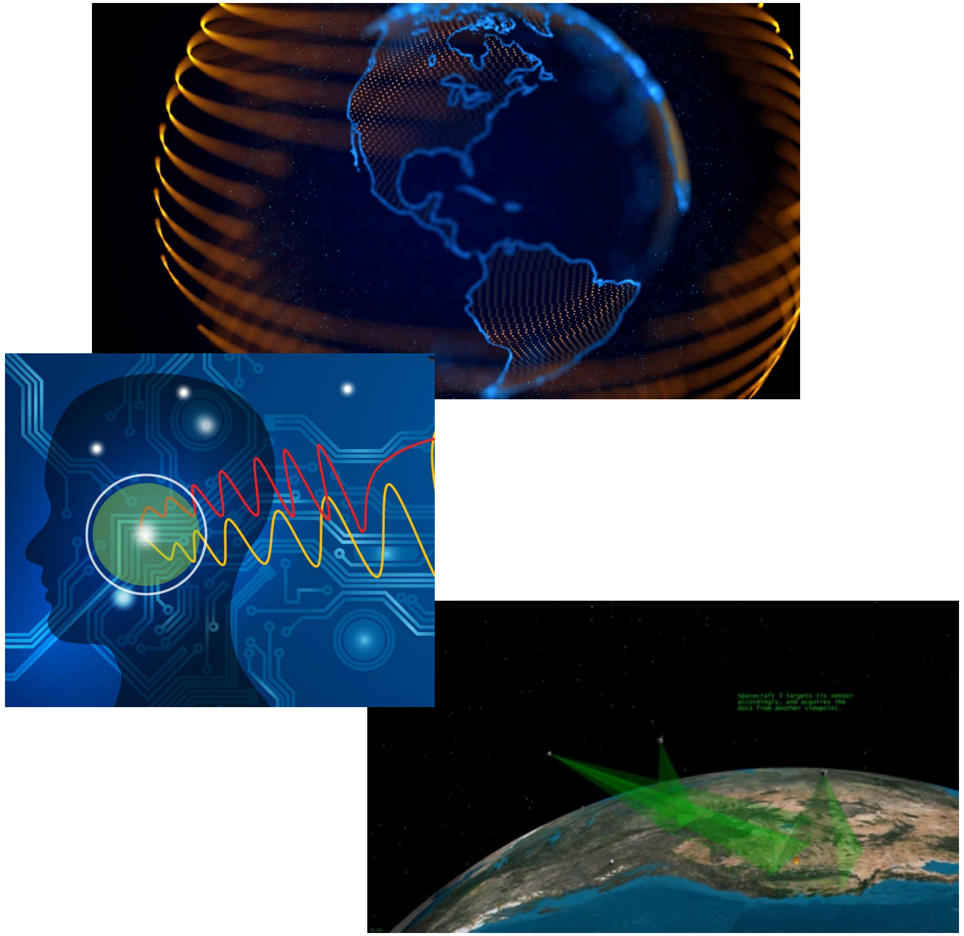


Machine Learning for NASA Advanced Information Systems

Jacqueline Le Moigne
NASA Earth Science Technology Office

November 14, 2022

AIST Objectives

Innovate in software and information systems technology that enable:

O1. New observation measurements and new observing systems design and operations through intelligent, timely, dynamic, and coordinated distributed sensing
=> New Observing Strategies (NOS)

O2. 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
=> Analytic Collaborative Frameworks (ACF)

O3. Developing 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. This thrust will provide technology for enabling near- and long-term science* and policy decisions
=> Earth System Digital Twins (ESDT)

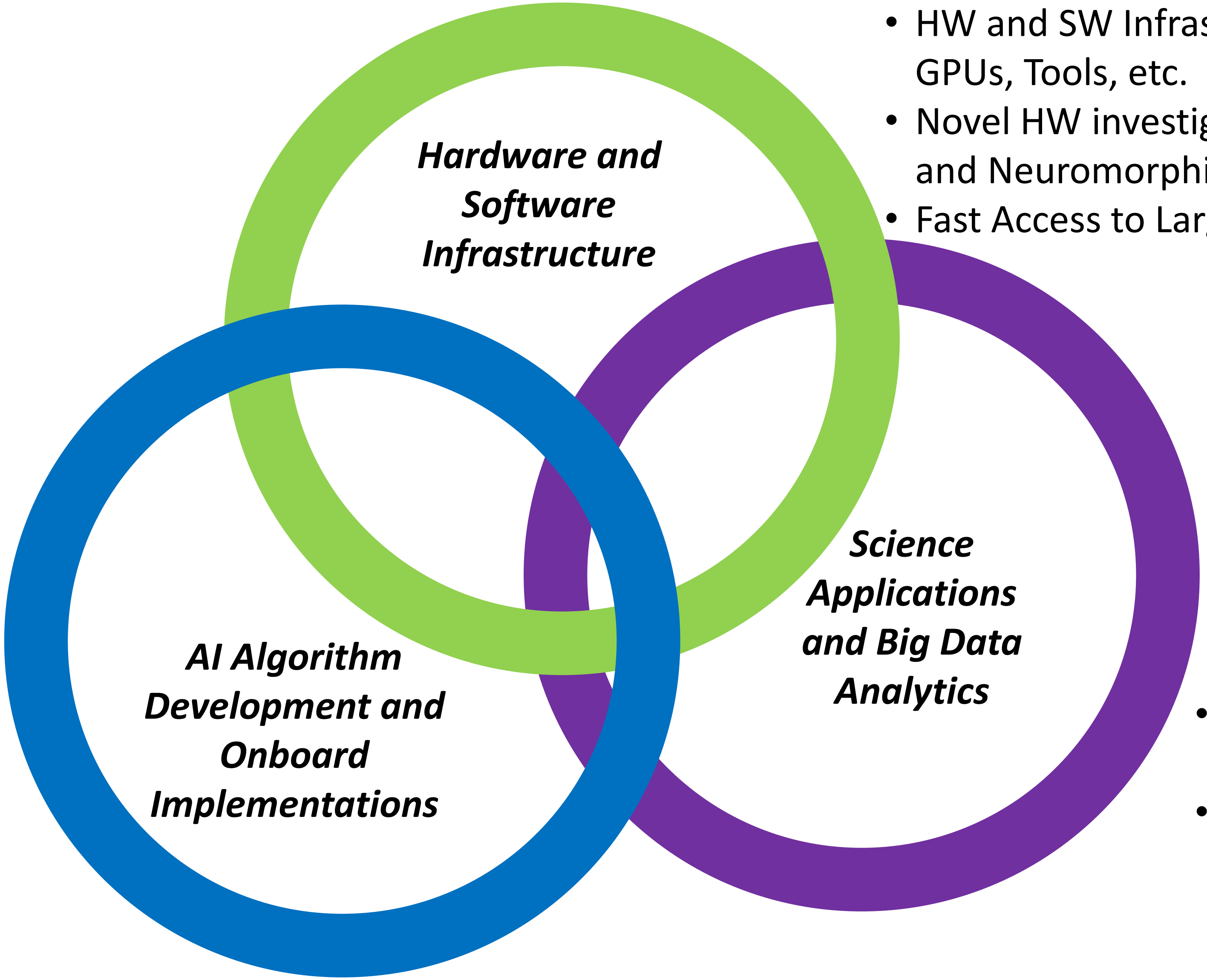
More generally, provide "Science Data Intelligence"

* "Science decisions" including planning for the acquisition of new measurements; the development of new models or science analysis; the integration of Earth observations in novel ways; applications to inform choices, support decisions, and guide actions for societal benefit; etc.

AI 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

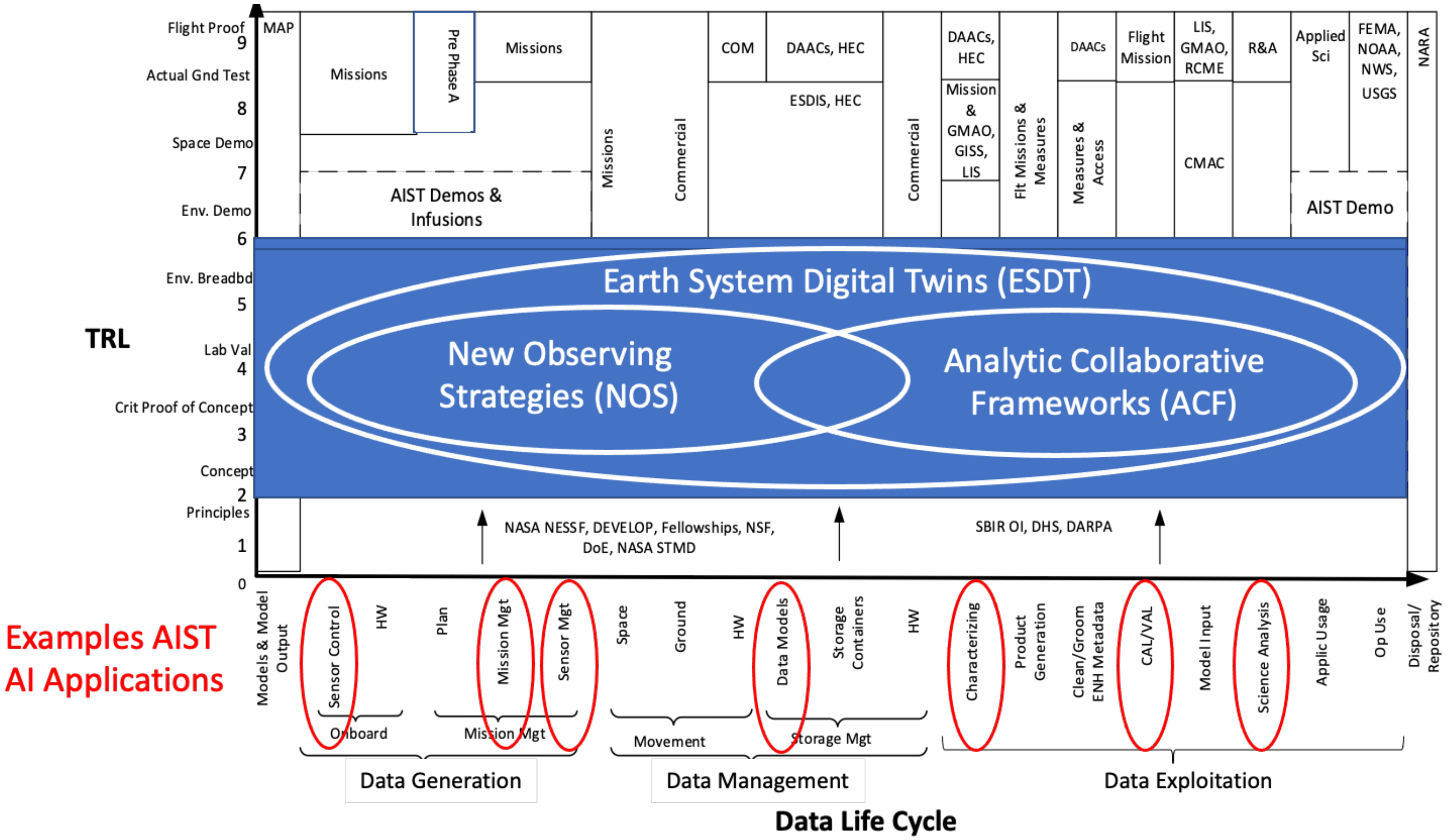
- AI Expertise
- Conceptual Software & Algorithm Development
- At the Edge/Onboard Implementations



- Science Applications and Data Analytics
- Algorithm Relevance and Validation

AIST Program Scope

NASA's Advanced Information Systems Technology (AIST) Program
identifies, develops, and supports adoption of advanced software and information systems, as well as
novel computer science technologies expected to be needed by the Earth Science Division in the 5-10-year timeframe



AI in AIST for Earth Science Applications

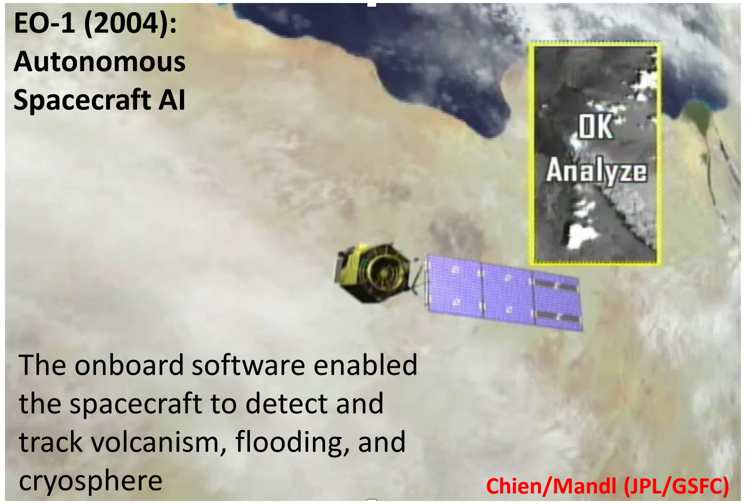
Two Main Areas

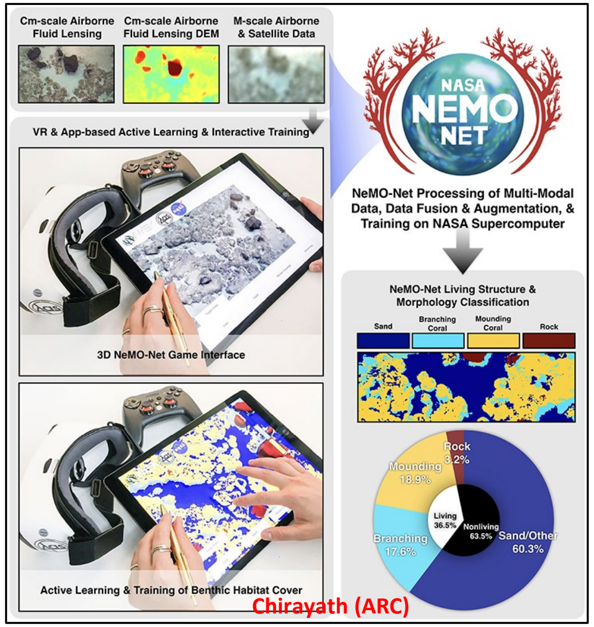
- **Improved Agile Observation Coordination and Mission Operations (Onboard or on the Ground) – *NOS***
  - At the edge data analysis
  - Semi-autonomy and autonomy for decision making
  - Anomaly and fault detection
  - Engineering Support for large constellations
  - Advanced Interoperability

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

- **Science Advancement – *ACF and ESDT***
  - 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

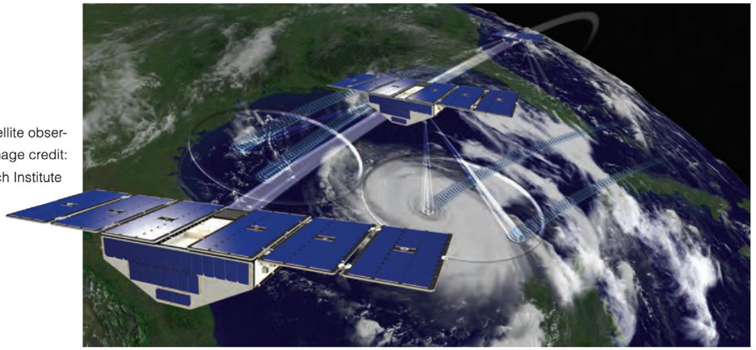
Technologies: Machine Learning/Deep Learning, Intelligent Search, Computer Vision, Data Fusion, Surrogate Modeling, Interactive Visualization & Analytics, Natural Language, …





Machine Learning for
New Observing Strategies (NOS)

NOS for Optimizing Measurements Design & Dynamically Capturing full Science Events

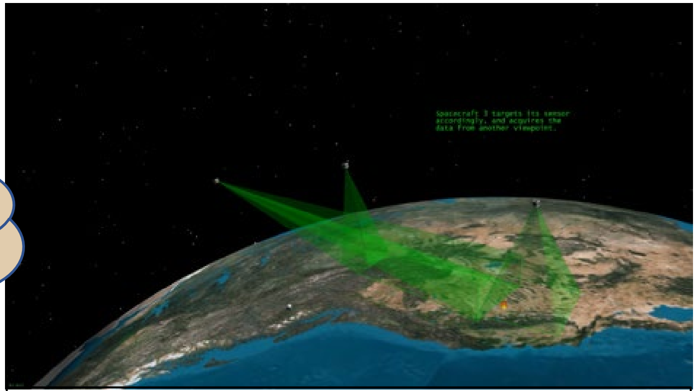


CYGNSS microsatellite observatories in orbit. Image credit: Southwest Research Institute

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

Provide complete picture of physical processes or natural phenomena

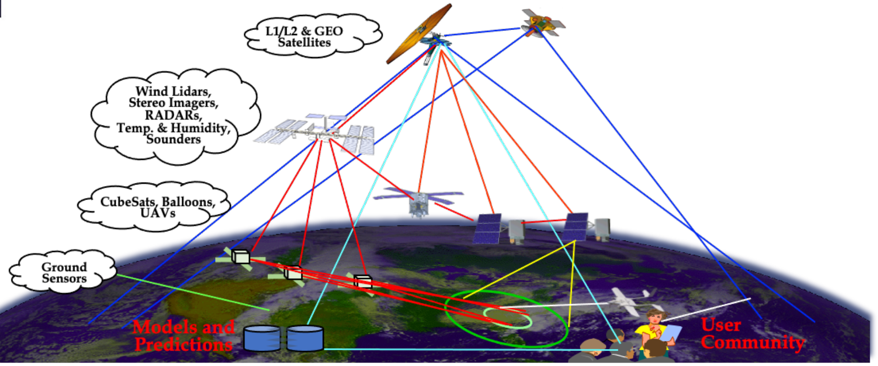
Increased understanding and predictability of dynamic events on Earth.



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:

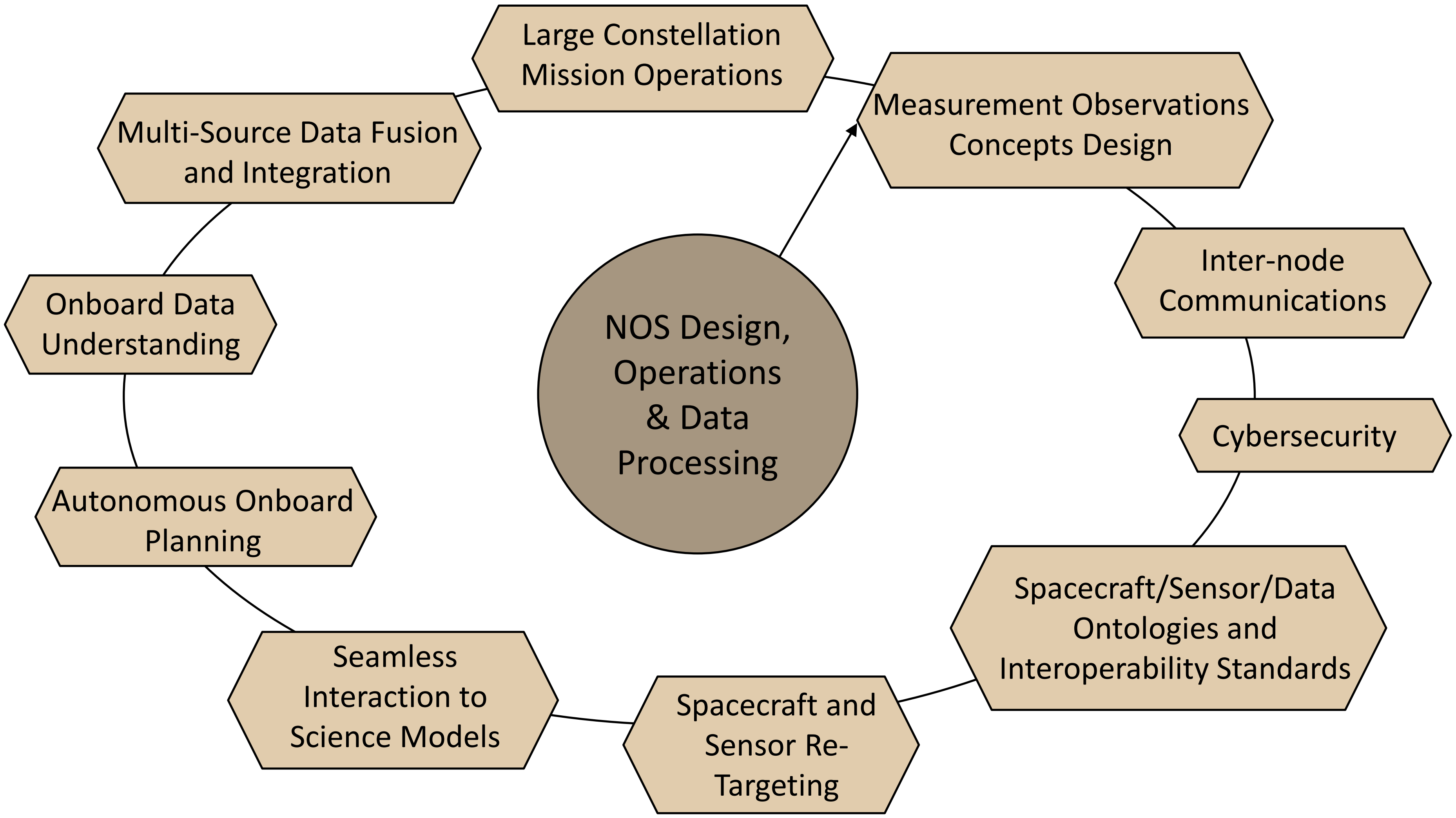
1. **Design and develop New Observing Concepts:**
   - From Decadal Survey or Model; **Various size spacecraft;** Systems of systems **(*Internet-of-Space*); Various organizations**
   - **Perform trades** on sensor number/type, spacecraft, orbits; resolutions; onboard vs. on-the-ground computing; inter-sensor 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

Technologies Needed for NOS

Some Examples of Capabilities Needed Onboard:

- Recognizing science events of interest
- Exchanging data inter-spacecraft
- Analyzing data for optimal science return
- Reconfiguring the spacecraft based on coordinated observations



AI in ESTO AIST Program
Novel Observing Strategies (NOS)

- **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) (Grogan/AIST-16 & QRS-19)
    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
  - SPCTOR: Sensing Policy Controller and OptimizeR (Moghaddam/AIST-18)
    Multi-sensor coordinated operations and integration for soil moisture, using ground-based and UAVs "Sensing Agents".
  - D-SHIELD: Distributed Spacecraft with Heuristic Intelligence to Enable Logistical Decisions (Nag/AIST-18)
    D-SHIELD is an operations design tool that will, for a given distributed space mission (DSM) architecture, plan re-orienting and operations of heterogeneous payloads, accounting for power/payload constraints while maximizing science value. It uses an iterative science observable simulator based on Observing System Simulation Experiments (OSSEs) adapted for real time planning and rapid mission design This project contributes to the New Observing Strategy (NOS) thrust area by developing an AI-based planning and scheduling-based DSM operations tool.

AI in ESTO AIST Program
Novel Observing Strategies (NOS)

- **AI for Coordinated Observation and Constellations**
  - Edge Intelligence for Hyperspectral Applications In Earth Science for New Observing Systems (Carr/AIST-21)
    Uses the SpaceCube processor and the TRL-5 SpaceCube Low-power Edge Artificial Intelligence Resilient Node (SC-LEARN) coprocessor [1] powered by Google Coral Edge Tensor Processing Units (TPUs) to implement two AI science use cases in hyperspectral remote sensing. The second (nighttime) application classifies artificial light sources after training against a catalog of lighting types. SC-LEARN was developed at GSFC for AI applications such as neural networks and is packaged in a small, low-power 1U CubeSat form factor. SC-LEARN will fly on STP-H9/SCENIC to the ISS with a Headwall Photonics HyperspecMV [3] hyperspectral imager.
  - Multi-Path Fusion Machine Learning for New Observing System Design and Operations (MacKinnon/AIST-21)
    Proposes to develop a system based on data fusion and multi-path neural network ML to aid in the design and operation of multi-sensor NOS concepts. Will build ML-enabled analytic tools and advanced computing environment capabilities for NOS workflows that utilize large amounts of diverse airborne and satellite observations. Using multiple neural networks working in parallel, it will first be demonstrated with a forest productivity use case, with fusion of lidar, spectrometry, satellite-derived climatology and ecosystem modeling providing insights into the driving environmental factors that influence productivity. Then will be used for sensitivity studies to guide sensor and mission requirements traceability.
  - 3D CHESS: Decentralized, Distributed, Dynamic and Context-Aware Heterogeneous Sensor Systems (Selva/AIST-21)
    Proposes to demonstrate proof of concept for a context-aware Earth observing sensor web consisting of a set of nodes with a knowledge base, heterogeneous sensors, edge computing, and autonomous decision-making capabilities. Context awareness refers to the nodes' ability to gather, exchange, and leverage contextual information to improve decision making and planning. Will demonstrate and characterize the technology in a multi-sensor in-land hydrologic and ecologic monitoring system performing 4 inter-dependent missions: studying non-perennial rivers and extreme water storage fluctuations in reservoirs and detecting and tracking ice jams and algal blooms.
  - Intelligent Long Endurance Observing System (Chandarana/AIST-21)
    Proposes the development of the Intelligent Long Endurance Observing System (ILEOS) to help scientists build plans to improve spatio-temporal resolution of climate-relevant gases by fusing coarse-grained sensor data from satellites and other sources and plan High-Altitude Long Endurance (HALE) UAS flights to obtain finer-grain data. ILEOS will also enable observations for longer periods and of environments not accessible through in-situ observations and field campaigns. 3 components: (1) the Target Generation Pipeline to identify candidate target scenes; (2) the Science Observation Planner using automated planning and scheduling technology to automatically generate a flight plan; and (3) a Scientists' User Interface.
  - An Intelligent Systems Approach to Measuring Surface Flow Velocities in River Channels (Legleiter/AIST-21)
    Will develop a New Observing Strategy (NOS) for measuring streamflow from a UAS using an intelligent system. Using the USGS/NASA UAS-based payload for measuring surface flow velocities in rivers (USGS & NASA), consisting of thermal/visible cameras, a laser range finder, and an embedded compute (integrated within a common software middleware), it will address both quality control during routine streamgaging operations by quantifying uncertainty, as well as autonomous route-finding during hazardous flood conditions using inter-sensor communications. Will be implemented for real-time processing onboard the platform.
  - A New Snow Observing Strategy in Support of Hydrological Science and Applications (Vuyovich/AIST-21)
    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.

SpaceBorne-2 Experiments/Chien (NASA JPL) –

ISS Onboard Processing Experiment

This project conducts 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*

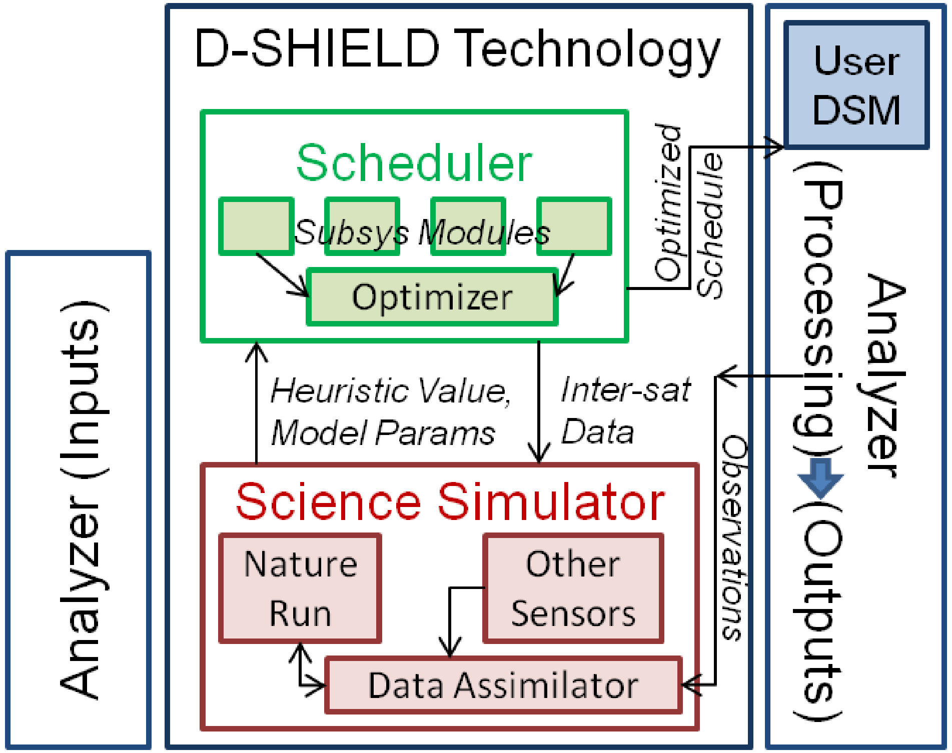
- **Re-Tasking Demonstration :** using onboard data analysis to create alerts
  - Use onboard data analysis to generate alerts, NOS- or SensorWeb-like, e.g,: task other assets
- **Live-Instrument Data Feed :** run experiments with data generated on ISS, e.g., ECOSTRESS, EMIT, OCO-3
  - Using both pre-uploaded 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

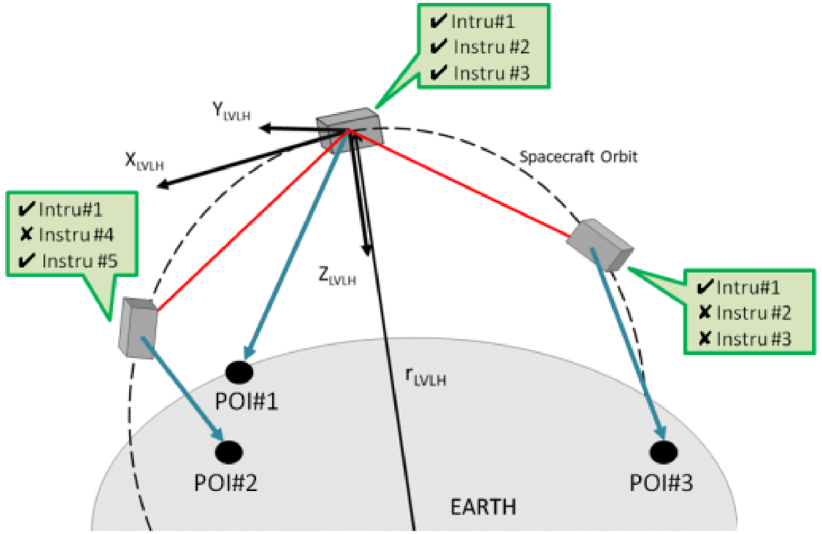
- **Radar Processing: l**everage NISAR and UAVSAR radar pipelines => data reduction, low-latency downlink
- **Thermal Infrared Processing :** experiment with TIR data from ECOSTRESS
  - Onboard pipeline: radiometric calibration, geolocation, land surface temperature, etc.
  - Applied Science Value: orders 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

AIST-18/Nag (NASA ARC) – *D-SHIELD: Distributed Spacecraft with Heuristic Intelligence to Enable Logistical Decisions*

D-SHIELD is an operations design tool that will, for a given distributed space mission (DSM) architecture, plan re-orienting and operations of heterogeneous payloads, accounting for power/payload constraints while maximizing science value. It uses an iterative science observable simulator based on Observing System Simulation Experiments (OSSEs) adapted for real time planning and rapid mission design This project contributes to the New Observing Strategy (NOS) thrust area by developing an AI-based planning and scheduling-based DSM operations tool.



D-SHIELD system diagram including data flows.



Cartoon of 3-satellite constellation with multiple instruments and D-SHIELD coordinated decisions

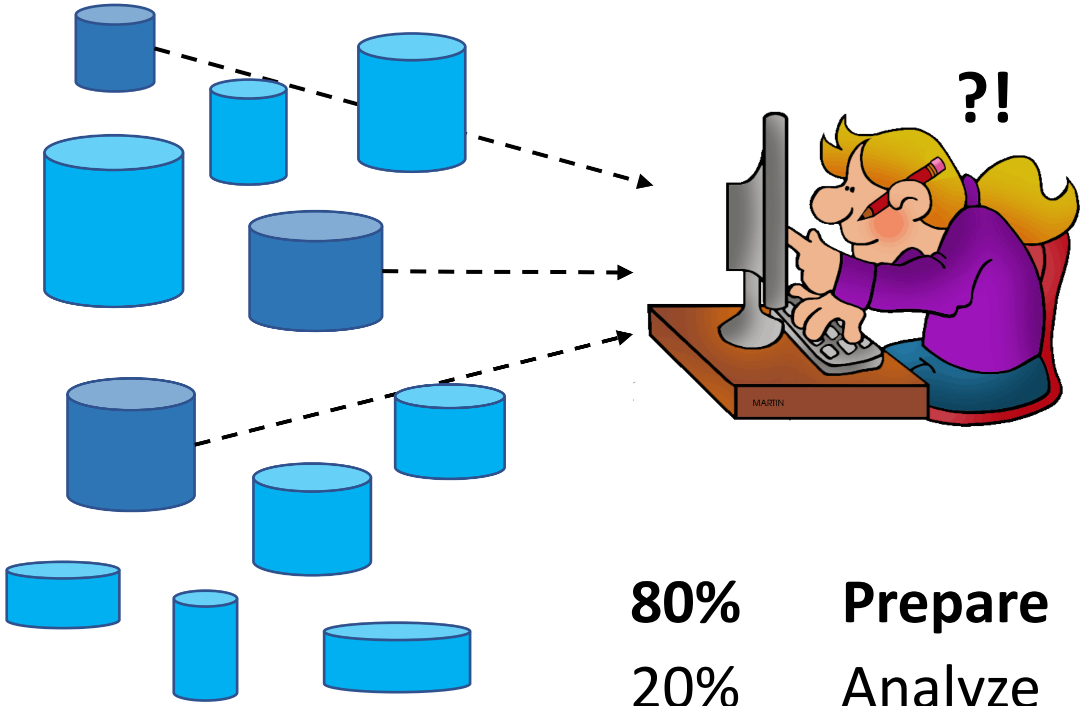
Machine Learning for Analytic Collaborative Frameworks (ACF)

From Archives to Analytic Frameworks
Focus on the Science User

Data Archives

Focus on data capture, storage, and management

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



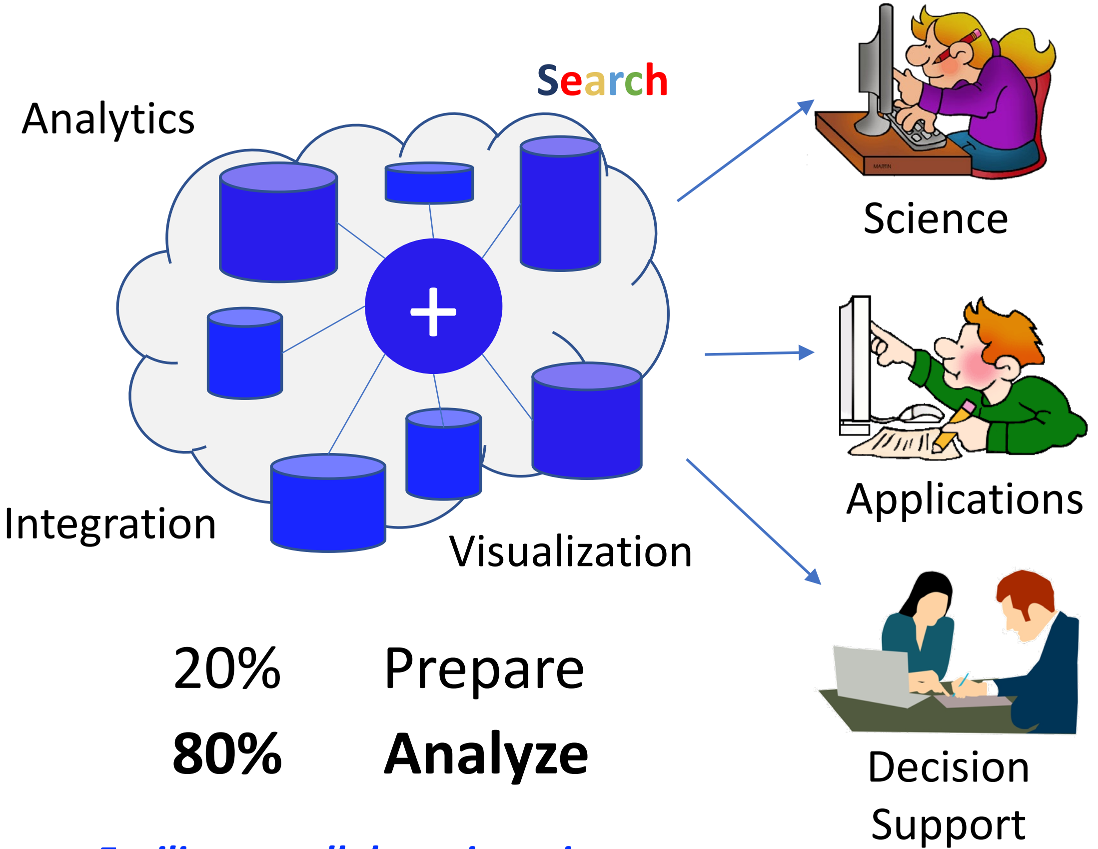
80% **Prepare**

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

Analytic Collaborative Frameworks (ACF)

Focus is on the Science User

Develop ACFs for new science domains:
- Manage large data volumes
- Manage wide variety of data types
- Manage frequent data updates (high data velocity)

AI CAPABILITIES:
- Machine Learning
- Deep Learning
- Data Services Discovery
- Uncertainty Quantification Methods

ADVANCED ANALYTICS:
- Data Accessibility
- Data Fusion
- Big Data Analytics
- Data Mining
- On-Demand Product Generation
- Data Operations Workflows
- Data Incorporation of Metadata, Provenance, Semantics, etc.

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

Computational Infrastructure
- Computing
  - Capacity
  - Capability
- Storage
- Communications

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

IMPROVED MODELING CAPABILITIES:
- Science Data Model Validation
- Software Architecture Frameworks
- Science Code Development & Reuse
- Modeling Systems
- Model Data Inter-Comparisons
- Custom Tools
- Forecasting/Prediction

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

AI in ESTO AIST Program
Analytics Collaborative Frameworks (ACF)

- **AI for Time Series and for Science Models**
  - 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.
  - Integration of Observations and Models into Machine Learning for Coastal Water Quality (Schollaert Uz/AIST-21)
    Coastal areas are impacted by population growth, development, aging infrastructure, and extreme weather events causing greater runoff from land. Monitoring water quality is an urgent societal need. Our AIST18 project demonstrated promising results with multispectral optical, medium spatial resolution satellite data trained using geophysical model variables within a machine learning (ML) architecture by extracting multi-source feature maps. This projects builds on that work by implementing higher resolution spectral and spatial data from commercial sources.

- **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.
  - SLICE: Semi-Supervised Learning from Images of a Changing Earth (Wilson/AIST-21)
    Proposes to investigate and characterize the efficacy of multiple Self- and Semi-Supervised Learning (SSL) techniques for representative image problems on Earth imagery, and then select the best for further infusion into mission and science workflows. Three top-level goals of the SLICE system are: (1) Establish the SLICE framework and platform on the AWS Cloud and supercomputing environments; (2) Investigate and characterize the accuracy of multiple SSL models (i.e. SimCLRv2, DINO, EsViT) on a variety of relevant remote sensing tasks with minimum labels (e.g., ocean phenomena); (3) Build and publish self- and semi-supervised learning models with a focus on the upper ocean small-scale processes in anticipation of several on-going and upcoming NASA missions (i.e. SWOT, WaCM, and PACE).

AI in ESTO AIST Program
Analytics Collaborative Frameworks (ACF)

- **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.
  - AMP – An Automated Metadata Pipeline (Huffer/AIST-18)
    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.

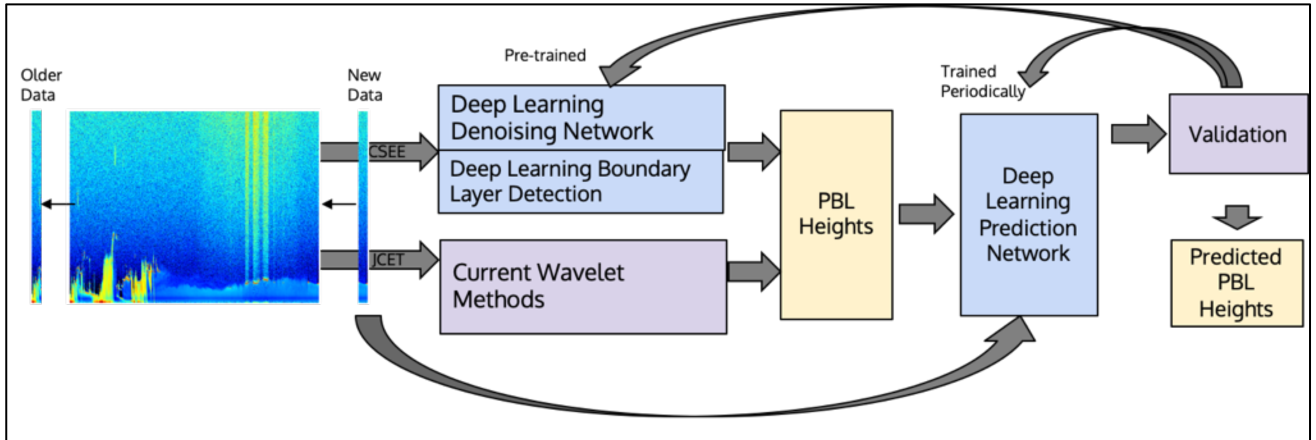
AI in ESTO AIST Program
Analytics Collaborative Frameworks (ACF)

- **AI for Pattern and Information Extraction (cont.)**
  - Innovative Geometric Deep Learning Models for Onboard Detection of Anomalous Events (Gel/AIST-21)
    The project proposes to enhance deep learning (DL) architectures for onboard learning with the most salient time-conditioned topological information of observed smoke plumes caused by wildfires. This enhancement will be achieved by integrating a fully trainable topological layer into time-aware deep neural networks (DNNs)
  - Knowledge Transfer for Robust GeoAI Across Space, Sensors and Time via Active Deep Learning (Prasad/AIST-21)
    The proect proposes to address harmonizing sensor-web data by developing a machine learning algorithmic framework and an associated open-source toolkit for robust analysis of multi-sensor remotely sensed data that advances emerging and promising ideas in deep learning.
  - Detection of artifacts and transients in Earth Science observing systems with machine learning (Bock/AIST-21)
    Proposes to create open-source software to provide a rich, interactive environment where machine learning (ML) models are used as collaborator to direct the attention of the human analyst to non-physical artifacts and real transient events that require interpretation. The proposed system will be realized through two coupled sub-systems: a novel "back-end" ML software called the Transient and Artifact Continuous Learning System (TACLS), and a significant upgrade to existing "front end" interactive MGViz user environment, originally designed to view displacement time series and their underlying metadata, to now interact and display layers of spatiotemporal information. Will use Scripp's archive of thousands of artifacts and transients.

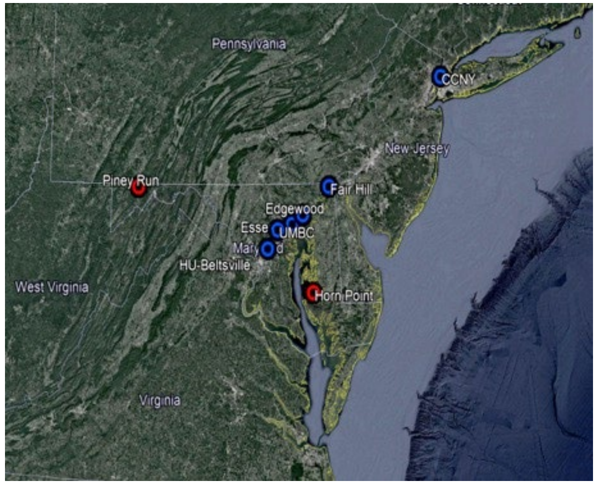
AIST-16/Halem (UMBC) –

Implement a Deep Learning Operational Ceilometer (LIDAR)-based Atmospheric Boundary Layer Height Product over Continental U.S.

This project prototypes 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

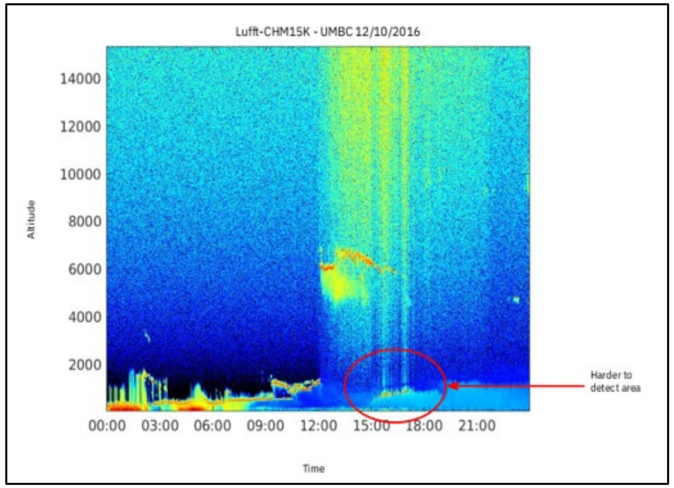
Radar Wind Profiler

- HU-Beltsville
- Horn Point Laboratory
- Piney Run

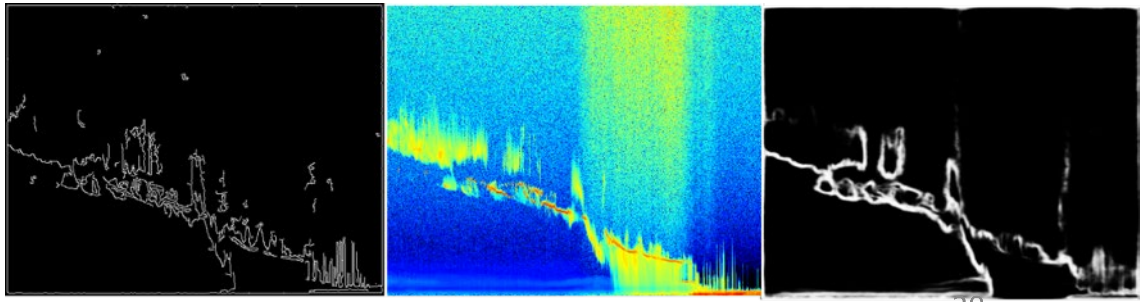
Current methods can be problematic under the following conditions:

- 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



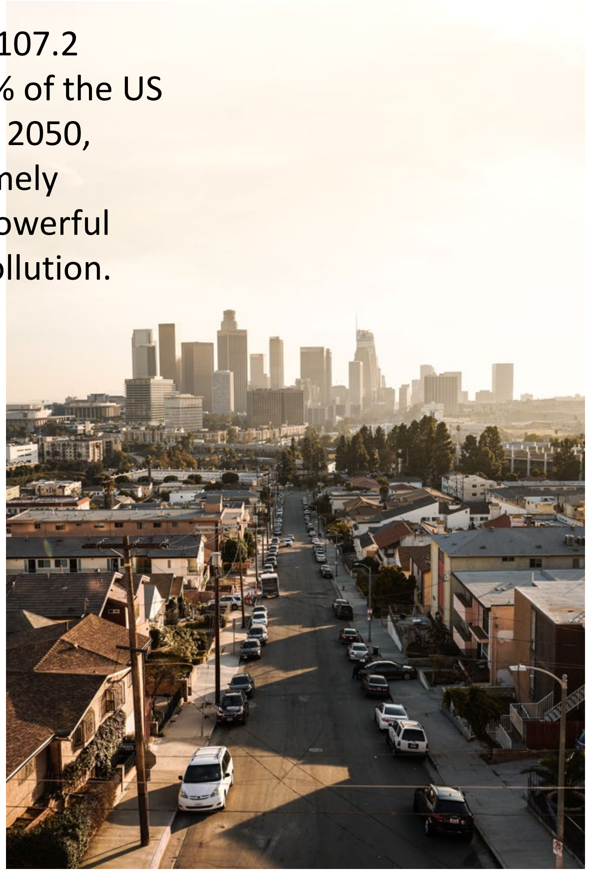
Canny Edge Detection | LIDAR Backscatter over 24 hours (Lufft-CHM15K Dec 6, 2016) | Deep Neural Network Edge Detection

AIST-18/Holm (City of Los Angeles) – *Predicting What We Breathe (PWWB)*

This project integrates ground-based, space-based data and machine learning to help cities predict air quality (AQ) and develop policies to mitigate air pollution. It will also improve the ability to forecast and mitigate air pollution events and may significantly reduce the negative environmental, economic and health impacts of these events in urban environments.

Air pollution is responsible for 4.5 million deaths and 107.2 million disability-adjusted-life-years globally. With 89% of the US population expected to live in urban environments by 2050, reducing air pollution-related loss of life will be extremely important. PWWB will provide policy makers with a powerful tool for reducing illnesses and deaths caused by air pollution.

- **New air quality initiative** led by the City of Los Angeles, in collaboration with OpenAQ (diverse fields including air quality research, data science, public health, city officials, and community activism.
- Using the **City of Los Angeles as a test case**
- Sister cities identified and recruited => air quality regime **similarities in cities around the world** being discovered
- Regional and international workshops to socialize the models, promote the open source, and gather requirements



PWWB website: https://airquality.lacity.org/

- Developed **5 predictive models** for analyzing air quality data based on ground-based sensors and satellite observations
- Provide predictions of PM2.5, NO2, CO, and O3
- **Ground-based air quality sensor data** from AQMIS and Port of LA datasets (arb.ca.gov)
- **Satellite observations** from NASA MODIS,ESA Sentinel-2; MERRA and MAIAC datasets; meteorological data including wind, temperature, humidity
- **Deep neural networks** based on the combination of **convolutional neural networks and recurrent neural networks** that discover both temporal and spatial patterns in the data

Machine Learning for
Earth System Digital Twins (ESDT)

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.

Earth System Digital Twins Components

Digital Replica . . .

An integrated picture of the past and current states of Earth systems.

Forecasting . . .

An integrated picture of how Earth systems will evolve in the future from the current state.

Impact Assessment . . .

An integrated picture of how Earth systems could evolve under different hypothetical what-if scenarios.

. . .

- **Continuous observations** of interacting Earth systems and human systems
- From many **disparate sources**
- Driving **inter-connected models**
- At many **physical and temporal scales**
- With fast, powerful and integrated **prediction, analysis** and **visualization** capabilities
- Using **Machine Learning, causality and uncertainty quantification**
- Running at **scale** in order to improve our **science** understanding of those systems, their **interactions and** their **applications**

Earth System Digital Twins Components

Digital Replica . . . What now?

An integrated picture of the past and current states of Earth systems.

Forecasting . . . What next?

An integrated picture of how Earth systems will evolve in the future from the current state.

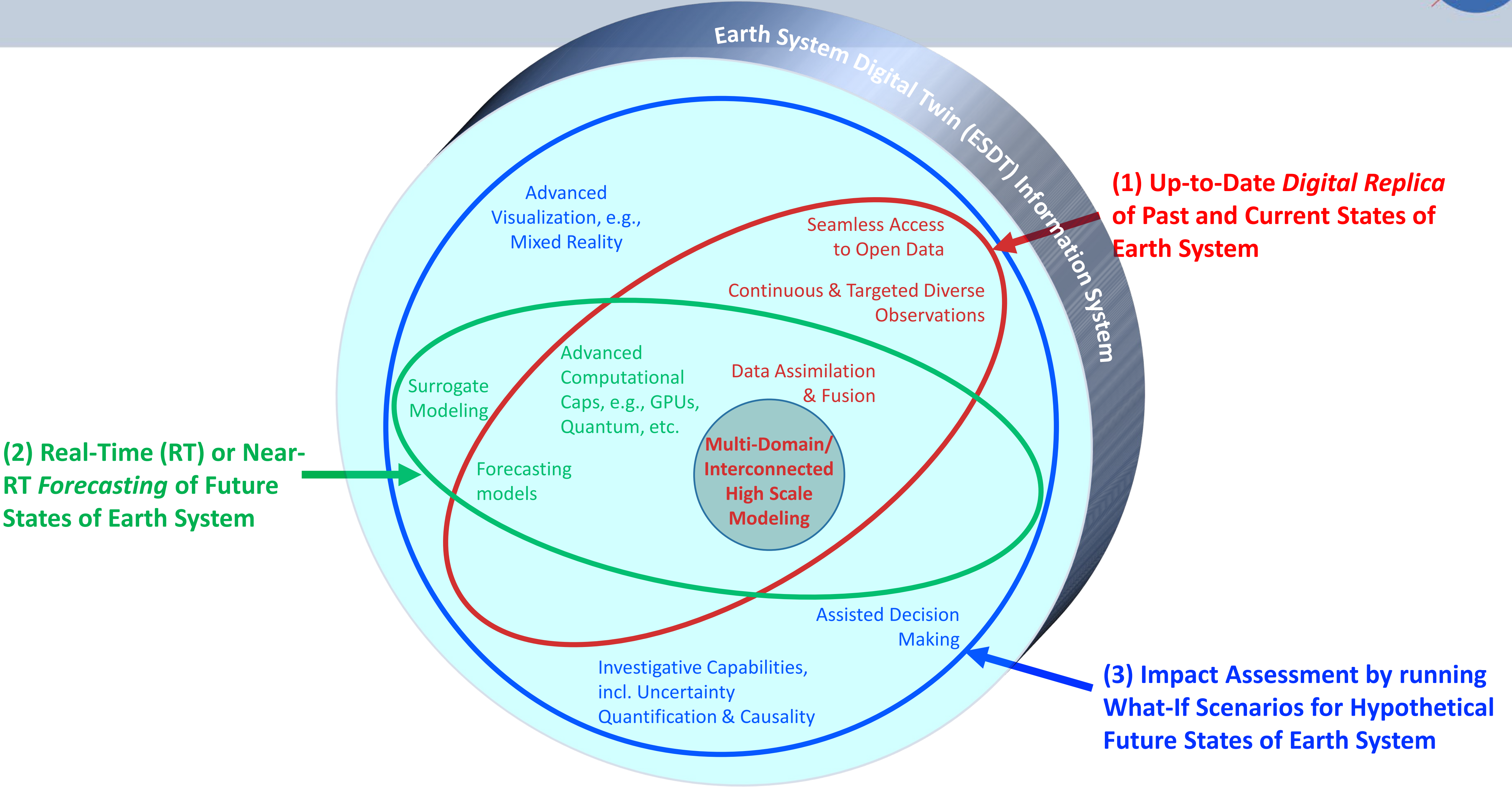
Impact Assessment . . . What if?

An integrated picture of how Earth systems could evolve under different hypothetical what-if scenarios.

. . .

- **Continuous observations** of interacting Earth systems and human systems
- From many **disparate sources**
- Driving **inter-connected models**
- At many **physical and temporal scales**
- With fast, powerful and integrated **prediction, analysis** and **visualization** capabilities
- Using **Machine Learning, causality and uncertainty quantification**
- Running at **scale** in order to improve our **science** understanding of those systems, their **interactions and** their **applications**

AIST ESDT Capabilities



AI in ESTO AIST
Earth System Digital Twins (ESDT)

- **AI for Data Assimilation**
  - Autonomous Moisture Continuum Sensing Network (Grubb/AIST21)
    Earth Science phenomena such as convective clouds, hurricanes and wildfire smoke plumes move with the 3-D flow field in a Lagrangian reference frame, and it is often difficult and unnatural to understand these phenomena with data on Eulerian grids. This project will develop a scientific exploration and analysis mixed augmented and virtual reality tool with integrated Lagrangian Dynamics (LD) to help scientists identify, track, and understand the evolution of Earth Science phenomena in the NASA GEOS model In addition, it will address the development of an Earth System Digital Twin that provides both a scientific discovery tool and a model analysis and improvement tool.
  - A Framework for Global Cloud Resolving OSSEs (Clune/ AIST-21)
    The main objective of this proposal is to enable global, cloud-resolving Observing System Simulation Experiments (OSSEs) by addressing key computational challenges that prevent existing technologies from scaling to the spatial resolutions that will be needed by the end of decade. Among the many beneficiaries, are NASA's Earth System Observatory (ESO) missions currently in development. To this end, we propose to design, implement, and deploy a user-friendly framework for performing global cloud-resolving Observing System Simulation Experiments (OSSEs).

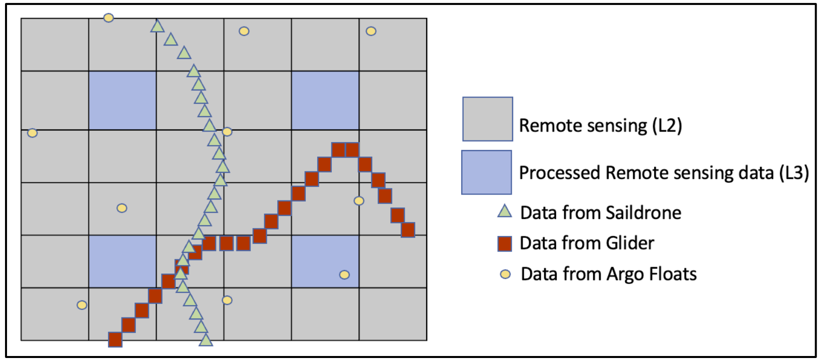
- **AI for Forecasting**
  - Advanced Phenology Information System (APIS) (Morisette/AIST-16)
    Ecological processes and uncertainty are evaluated by fitting a Bayesian hierarchical model to annually oscillating time series of vegetation indices, with the R package "greta", which utilizes TensorFlow 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.
  - 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.
  - Predicting What We Breathe: Using Machine Learning to Understand Urban Air Quality (Holm/AIST-18)
    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.
  - Development of a next-generation ensemble prediction system for atmospheric composition (Keller/AIST-21)
    We propose to develop a next-generation modeling framework for the real-time simulation of reactive gases and aerosols in the atmosphere. The core innovations of this project are (a) the deployment of computationally efficient parameterizations of atmospheric chemistry and transport and (b) the development of generative models based on machine learning (ML) to predict model uncertainties. A key aspect of the proposed system is that the original numerical model and the accelerated models can be used in tandem. This way, the full physics model can be deployed for the main analysis stream, and the accelerated system is used to improve overall analytic and predictive power during forecast and data assimilation. This minimizes the impact of compounding errors that can arise from the use of ML models alone.

AI in ESTO AIST
Earth System Digital Twins (ESDT)

- **AI for Digital Twin Prototypes**
  - Towards a NU-WRF based Mega Wildfire Digital Twin: Smoke Transport Impact Scenarios on Air Quality, Cardiopulmonary Disease and Regional Deforestation (Halem/Firetech-22)
    The project will develop, implement and evaluate a regional Wildfire Digital Twin (WDT) model for the simulation of fire, its spread and smoke impacts on air quality, regional deforestation and cardiopulmonary disease on local and distant communities. WDT will embed within the NASA Unified Weather Research Forecasting (NUWRF) model, a data driven, chemical based, interactive atmospheric fire spread resolving module (SFIRE) and test the performance and long-term stability of deep, dense, physics aware machine learning emulations of the HRRR microphysics compatible with a rapid refresh hourly prediction.
  - A Prototype Digital Twin of Air-Sea Interactions (Gray/AIST-21)
    Proposes to develop hybrid physics-informed AI model that ingests several existing flux estimates and observation data products and train against simultaneous ocean-atmosphere data from Saildrones. This will ascertain uncertainty of existing flux measurements and optimize combination of near-real-time existing flux data and observational data => This represents the first step towards a Digital Twin for the Planetary Boundary Layer.
  - Terrestrial Environmental Rapid-Replicating Assimilation Hydrometerology (TERRAHydro) System: A Machine Learning-coupled Water, Energy, and Vegetation Terrestrial Earth System Digital Twin (Pelissier/AIST-21)
    Proposes to develop a Terrestrial Earth System Digital Twin (TESDT) that couples state-of-the-art ML with NASA (and other) EO data. It will combine the best ML hydrology models with capabilities for uncertainty quantification and data assimilation to provide ensemble & probabilistic forecasting, sensitivity analyses, and counterfactual "what if'' experiments.

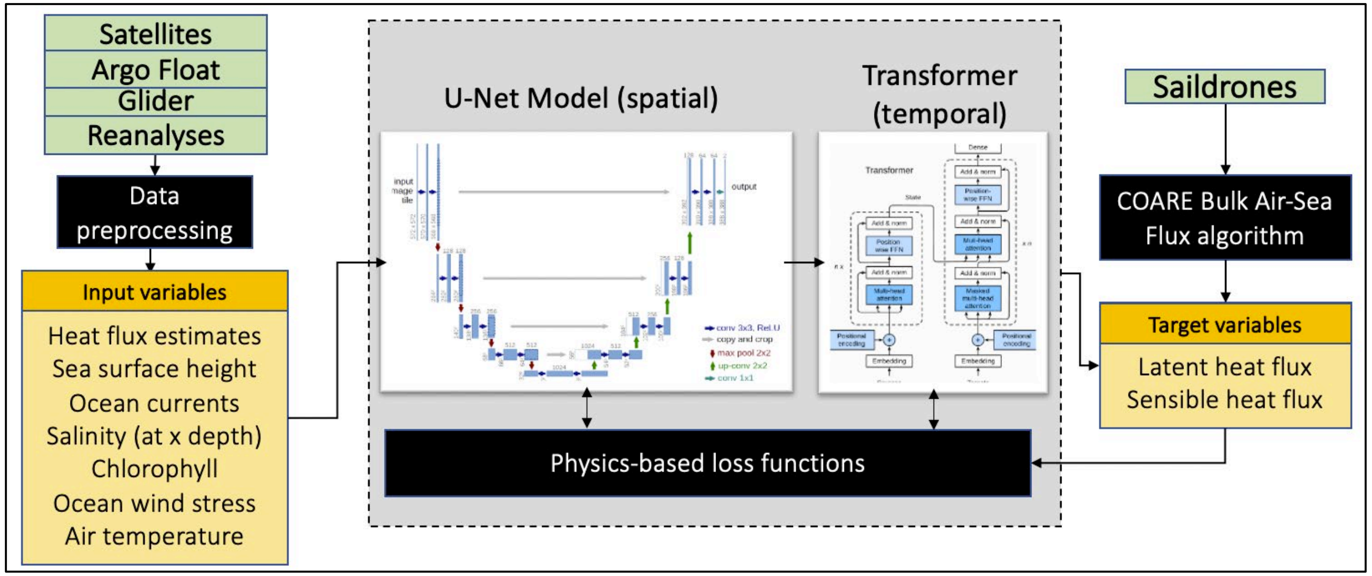
AIST-21/Gray (Univ. of Washington) –
A Prototype Digital Twin of Air-Sea Interactions

This project proposes to develop hybrid physics-informed AI model that ingests several existing flux estimates and observation data products and train against simultaneous ocean-atmosphere data from Saildrones. This will ascertain uncertainty of existing flux measurements and optimize combination of near-real-time existing flux data and observational data. This represents the first step towards a Digital Twin for the Planetary Boundary Layer.



The core technical challenge of this project is the development of the fully differentiable **hybrid model** to capture the boundary layer air-sea interactions. This model will be capable of realizing the non-linear interactions between atmospheric and oceanic surface variables such as sea surface temperature, wind speed and direction, and ocean currents, and the four different components of the heat fluxes. The model development task comprises two data-driven sub-components with physics-informed loss functions and constraints as described below.

- **Data-driven component:** A graph sampling method will be used to query multidimensional input-output data pairs. This data will be modeled using a U-net architecture that is specifically designed to handle spatial data.
- **Physics-informed constraints:** Constraining the model predictions during learning is a critical aspect of the model development to infuse domain knowledge. Constraints will be enforced in two ways:
  1. With ***soft enforceable physics constraints*** in the form of Lagrangian multiplier regularization losses. For example, changes in sea surface fluxes must be balanced by changes in atmospheric sensible heat and energy transport by the circulation.
  2. With ***hard constraints*** using periodic padding on the U-net architecture's direct neighbor layer. The data obtained from the in-situ observation platforms (Saildrones) will be used to validate and calibrate the data-driven model.

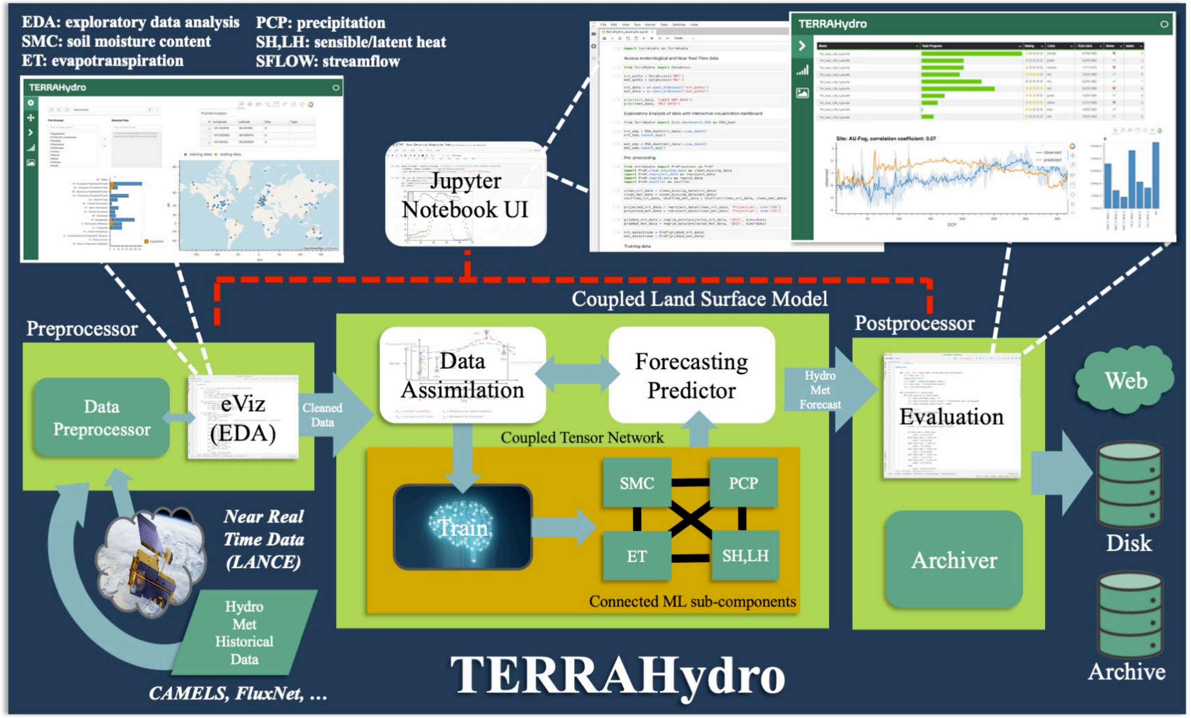


AIST-21/Pelissier (SSAI) – *Terrestrial Environmental Rapid-Replicating Assimilation Hydrometerology (TERRAHydro) System*

This project proposes to develop a Terrestrial Earth System Digital Twin (TESDT) that couples state-of-the-art ML with NASA (and other) EO data. It will combine the best ML hydrology models with capabilities for uncertainty quantification and data assimilation to provide ensemble & probabilistic forecasting, sensitivity analyses, and counterfactual "what if'' experiments.

The modeling framework will include several major components:

- Data ingestion and pre-processing pipelines that are model agnostic and allow for puzzle-piece integration with ML models.
- Individual ML models of specific land surface variables.
- Mechanisms for coupling individual ML models into a full terrestrial systems model.
- Support for integrating process-based modeling components.
- Support for real-time data assimilation.
- Tools for training, evaluation, scenario analysis, and benchmarking.



Various ML models have been developed and trained for many of the most critical terrestrial states and fluxes, e.g., for Soil Moisture, Streamflow, Surface Heat Fluxes, Net Ecosystem Exchange, and Snowcover. The goal of the project is to support multiple coupling approaches (e.g., direct coupling, shared model structure or PDE-based learning) and to provide the tools and interfaces that allow researchers to develop and test new coupling strategies. This is possible in an effectively general way because of the modular nature of tensor networks.

Machine Learning for Early-Stage Technology (EST)

AI in ESTO AIST
Early-Stage Technology (EST)

- **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.

- **AI for Smart Sensor**
  - Sensor-in-the-Loop Testbed to Enable Versatile/Intelligent/Dynamic Earth Observation (VIDEO) (Blackwell/AIST-21)
    Develops a methodology and test approach for a scene measured by a sensor to be able to configure the sensor in real-time during the scene measurement. Will significantly improve the resolution of the retrieved atmospheric fields in regions in which that improvement is most beneficial, while conserving resources in other regions. Includes two components: (1) Radiometric Scene Generator (RSG) using advanced metamaterial and its associated control software; (2) Intelligent processing and configuration software using feature detection and ML, running onboard the sensor, to detect and react to changes by dynamically optimizing the sensor response functions.

Conclusion

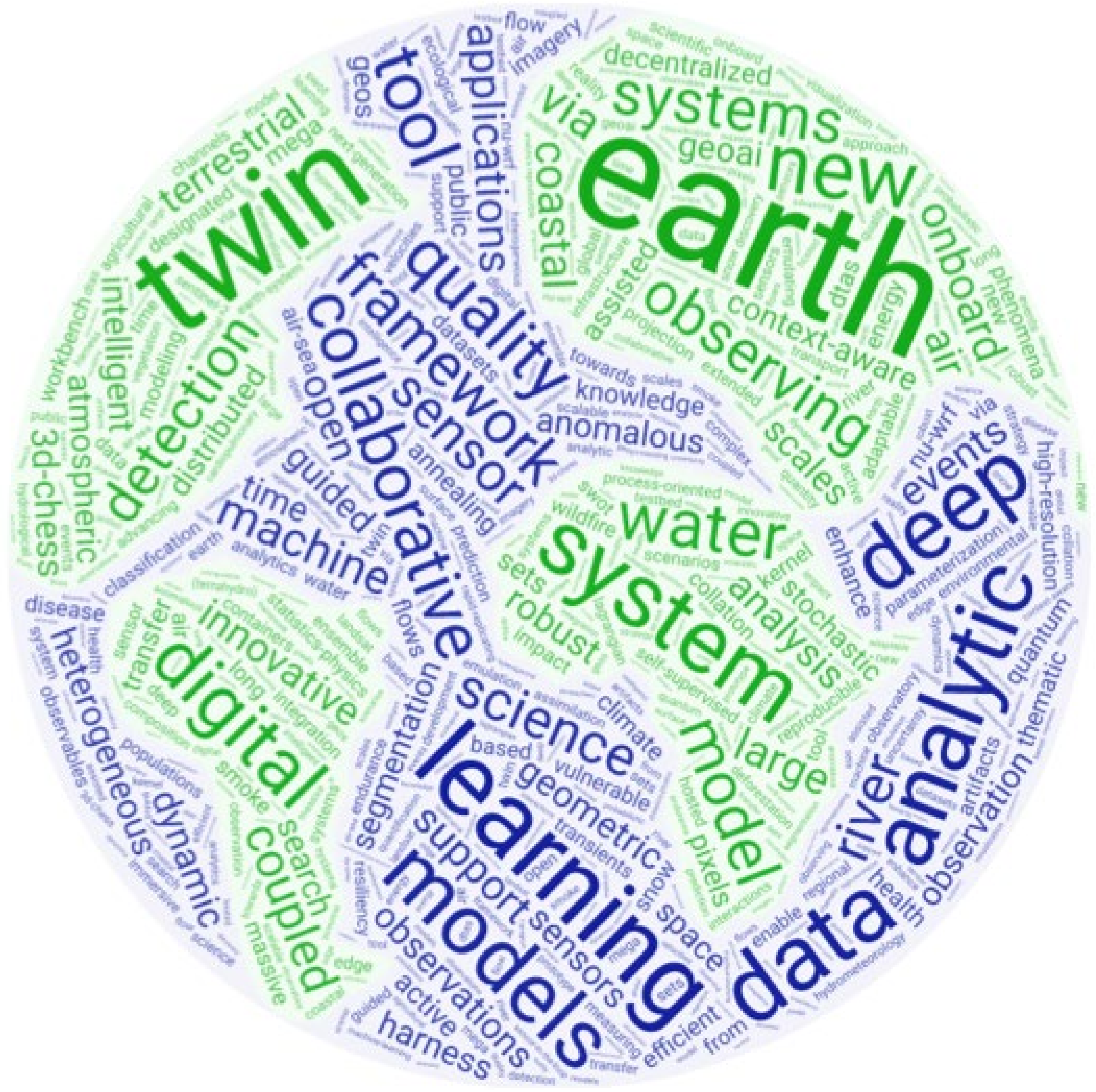
In Summary

- Machine Learning (and more generally AI) can be used in many steps of the Data Lifecycle
  - For optimizing and coordinating the acquisition of new measurements and the design of future observing systems
  - For the fusion, mining and analysis and ever-increasing amounts of remote sensing data as well as diverse data (including non-traditional unstructured data such as from IoT)
  - For speeding up modeling and prediction capabilities

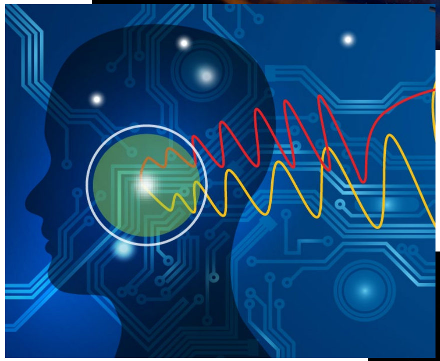
- In particular, for Earth System Digital Twins, Machine Learning and other AI techniques are important tools for all 3 components:
  - Digital Replica: for fusion, data assimilation and acquisition of relevant and timely observations and for model-coupling
  - Forecasting: for surrogate modeling, fast predictions
  - Impact Assessment: for visualization and what-if investigations

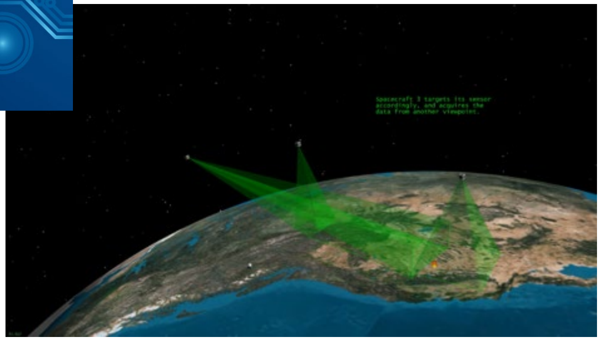
- Challenges are:
  - Validation; Accuracy/Uncertainty Quantification; Explainability
  - Hybrid Physics-informed models
  - Transfer Learning

Questions?



AIST-21 Word Cloud of Selected Proposals Titles



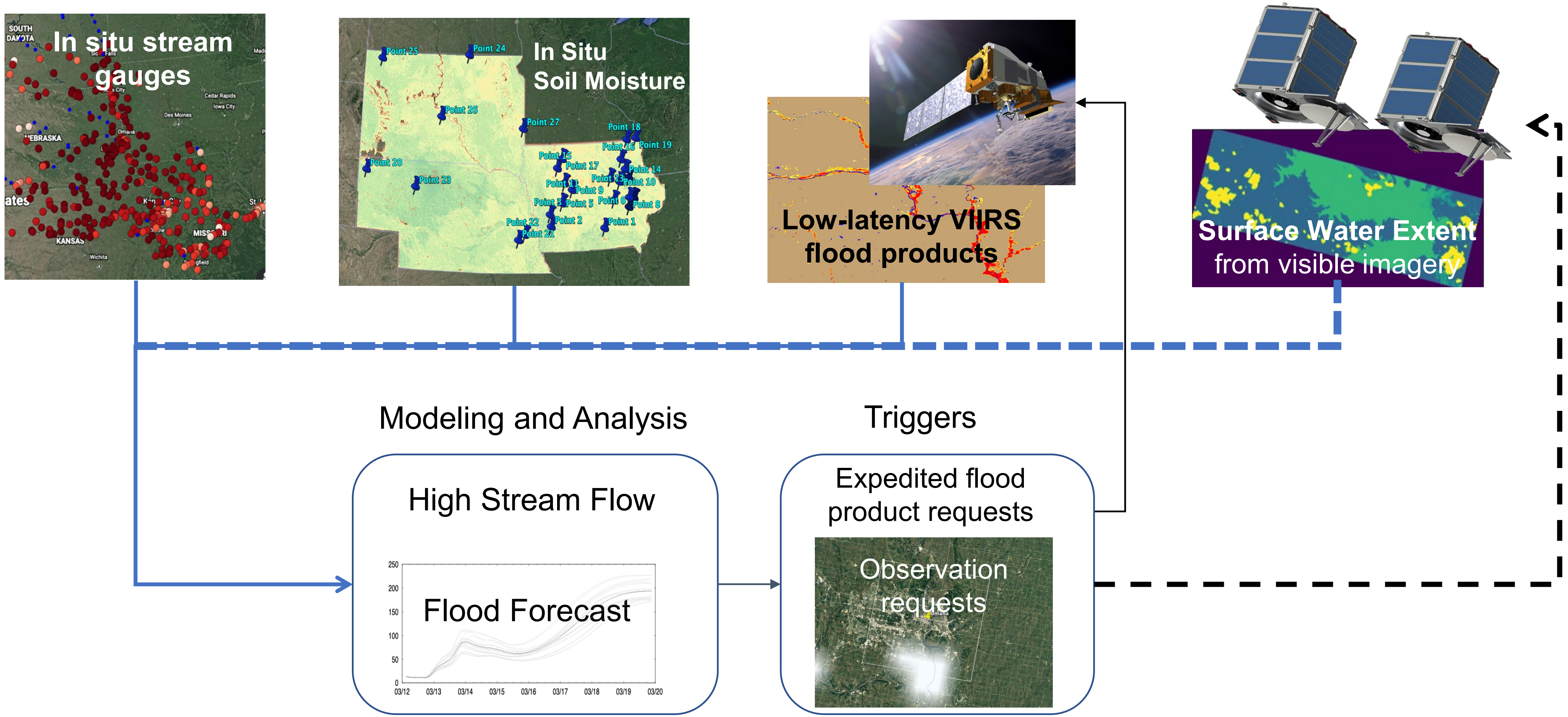


Back-up

Other Projects Details

NOS-Testbed Hydrology Demonstrations

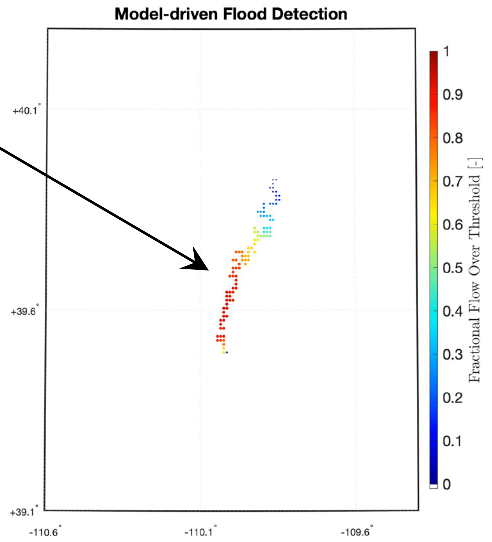
March 2021 – Historical Nebraska Flood + Live Mid-West Flood



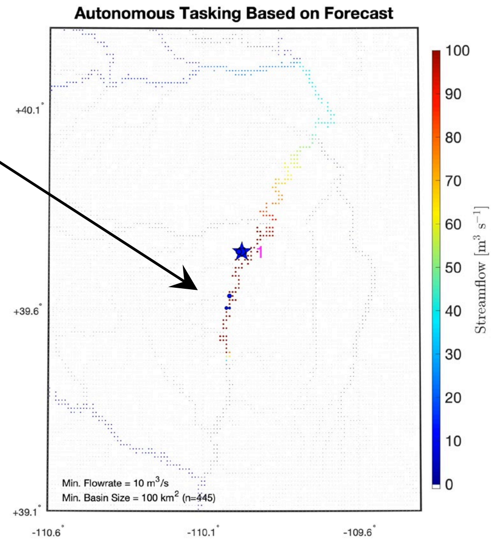
NOS-L: Autonomous, Model-driven Tasking of Flood Events

Successful NOS-L end-to-end live demonstration

flood wave exceeding climatological threshold



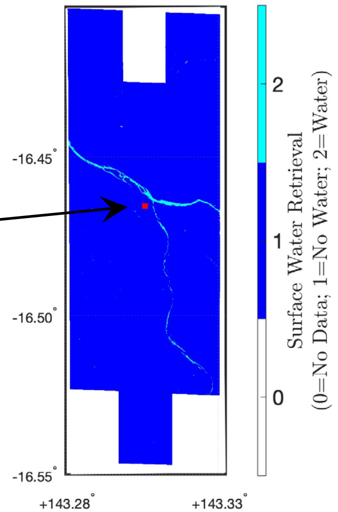
model-driven tasking of commercial satellites for observation collection (different marker sizes represent different forecast lead times)



Automated tasking **enables timely, coincident observation collection** to improve model predictions via **data assimilation**

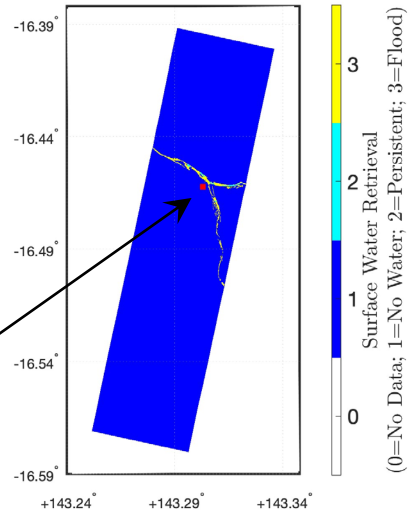
Planet

surface water extent retrieval via **VIS/NIR radiometry** (targeted location in red)



Capella Space

surface water extent retrieval via **X-band SAR** (targeted location in red)



AIST-18/Beck (U. Alabama, Huntsville) –
Cloud-Based Analytic Framework for Precipitation Research

Leverages 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

Cloud-based Analytic Framework for Precipitation Research

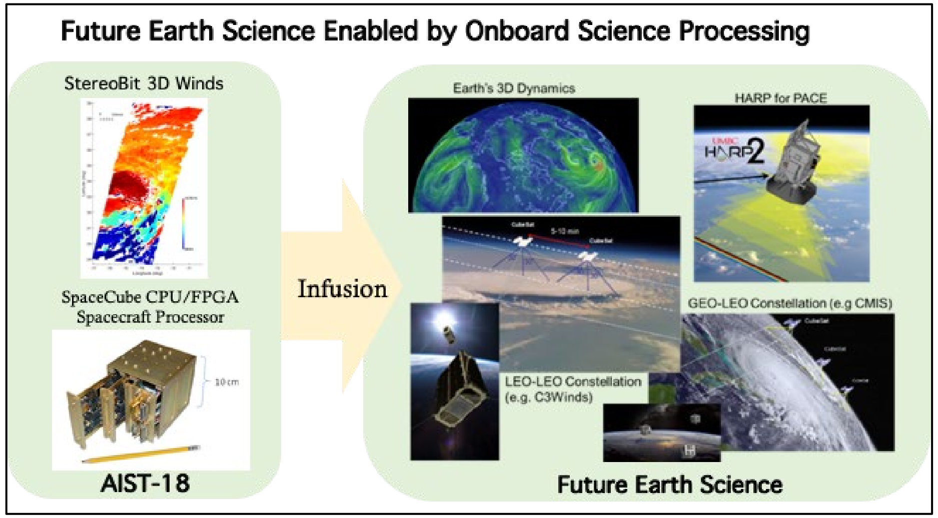
CAPRi

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

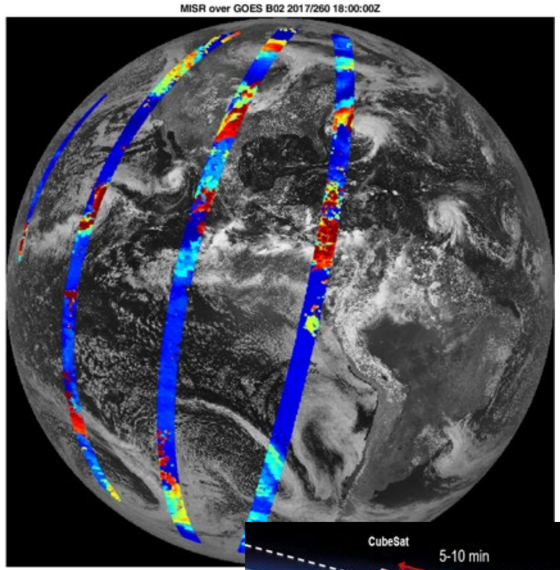
1. Providing new solutions for **real-time querying** of large data sets in a serverless environment;
2. Developing new methods for **generating Deep Learning Training data** on the fly; and
3. Providing an **easy-to-use user interface** for analysis and visualization of the data.

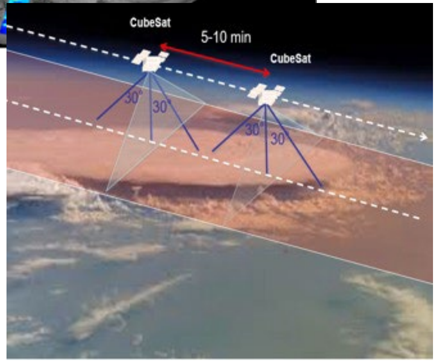
AIST-18/Carr (Carr Astro) – *StereoBit: Advanced Onboard Science Data Processing to Enable Future Earth Science*

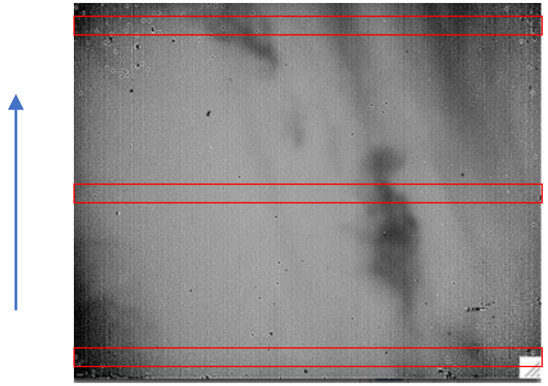
This investigation will demonstrate higher-level onboard science data processing for more intelligent SmallSats and CubeSats to enable future Earth science missions and Earth observing constellations. Low-cost SmallSat architectures generally suffer from downlink bottlenecks and often result in lower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

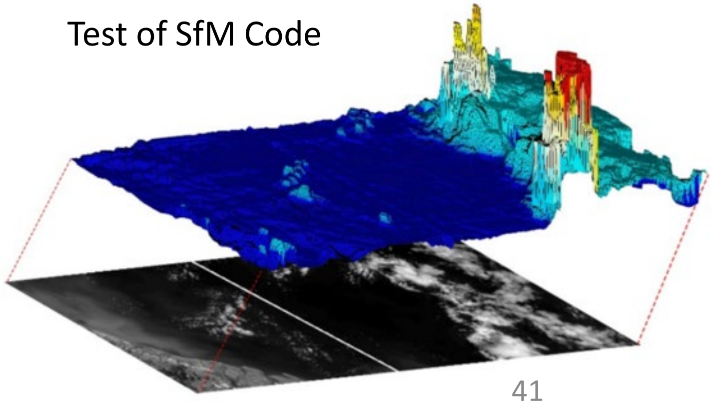






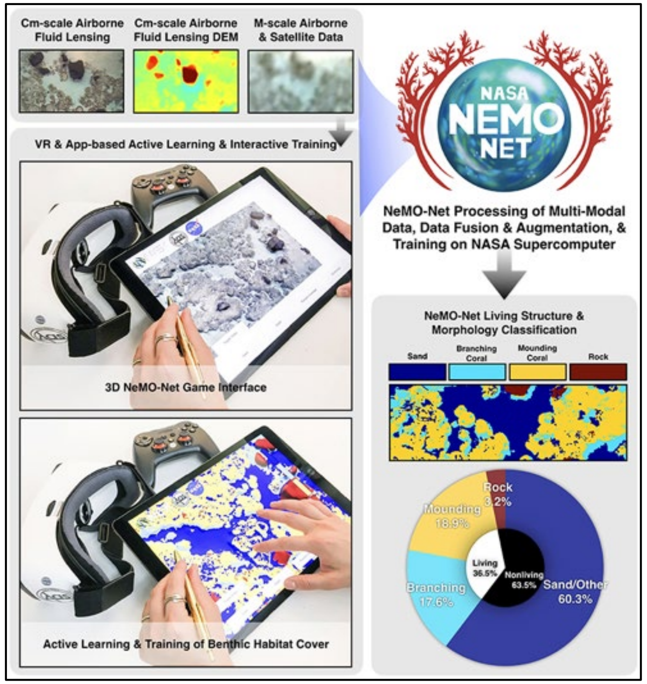
Early CTI Cloud Picture

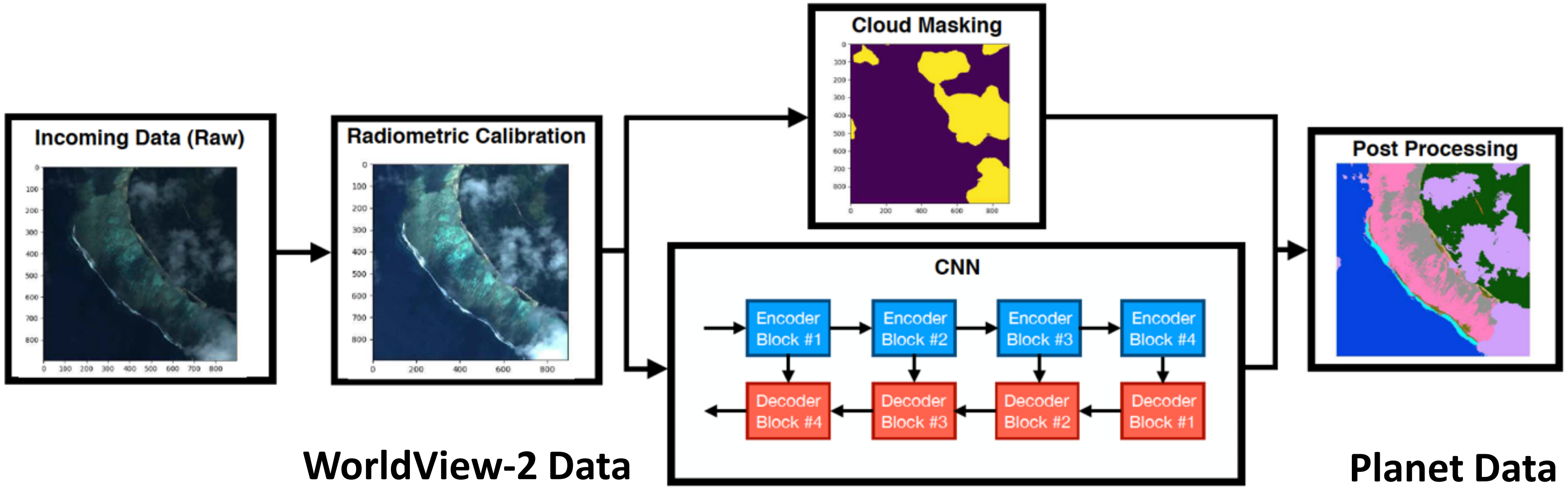
Test of SfM Code



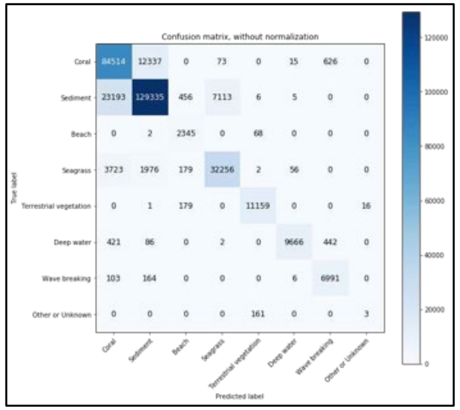
AIST-16/Chirayath (NASA ARC) –
NeMO-Net - The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment

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.



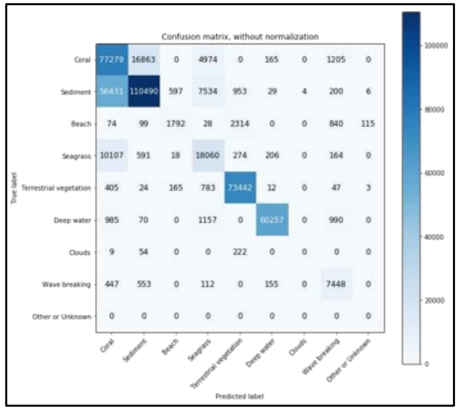


WorldView-2 Data



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

Planet Data

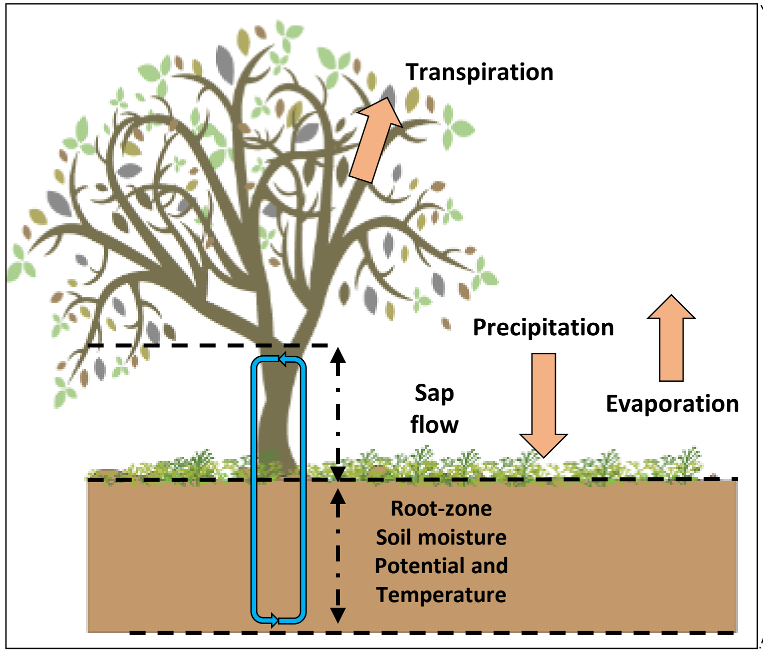


Total Accuracy: 76.0%
Accuracy amongst coral, sediment, seagrass classes: 68.1%

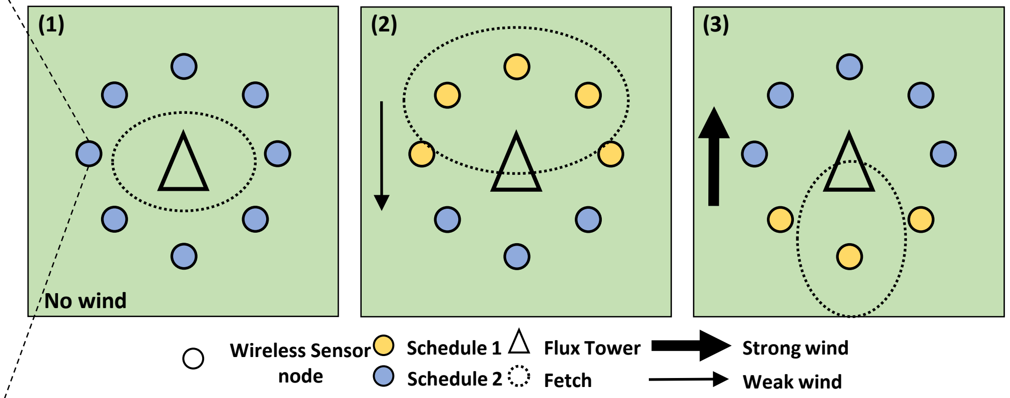
AIST-16/Entekhabi & Moghaddam (MIT & USC) –

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 .



Spatially Adaptive Sampling



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.

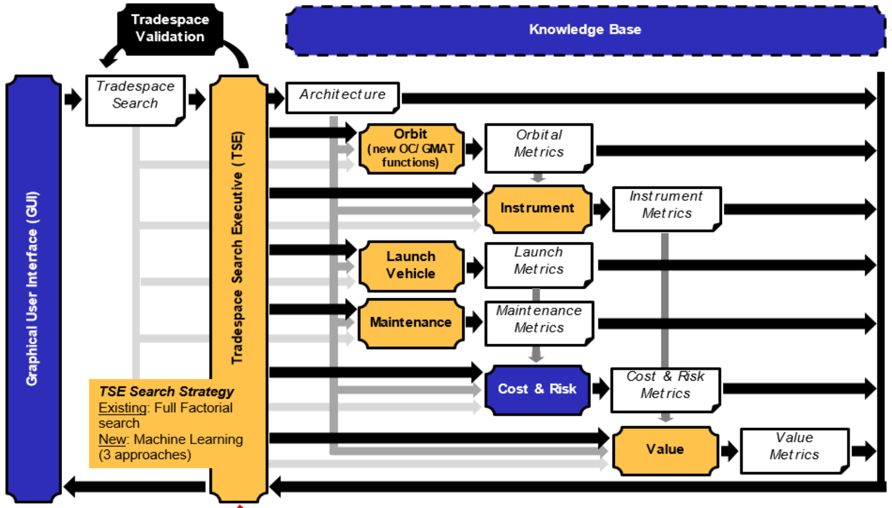
- ✓ Information Gain vs. Energy Consumption optimization → present as Pareto Fronts
- ✓ An autoregressive ML will have superior performance $\theta(t) = f(\theta(t-1)) + g(X(t))$
- ✓ Simple Policies can achieve superior RMSE performance with less energy consumption

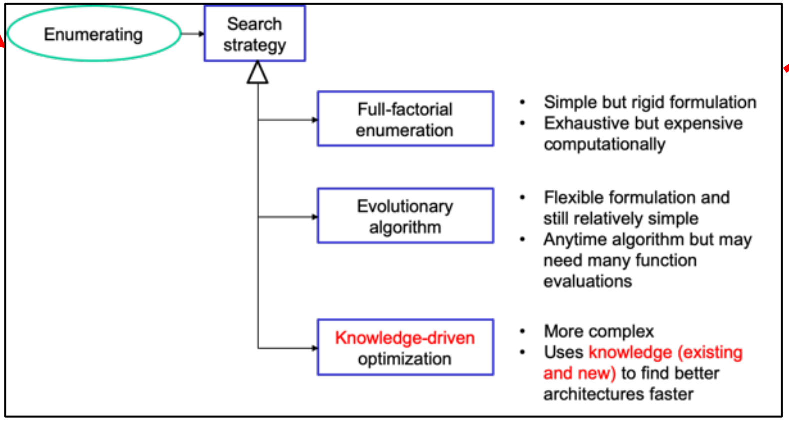
- SoilSCAPE Plan → Satellites Cal/Val
  - SMAP Cal/Val: Deployed 1 site at the Cary Institute of Ecosystem Studies (Millbrook, NY)
  - SoilSCAPE team (via. Co-I Moghaddam) collaborating with CYGNSS to provide *in situ* soil moisture for cal/val activities
  - Established a cal/val infrastructure for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

SoilSCAPE installation for CYGNSS Cal/Val

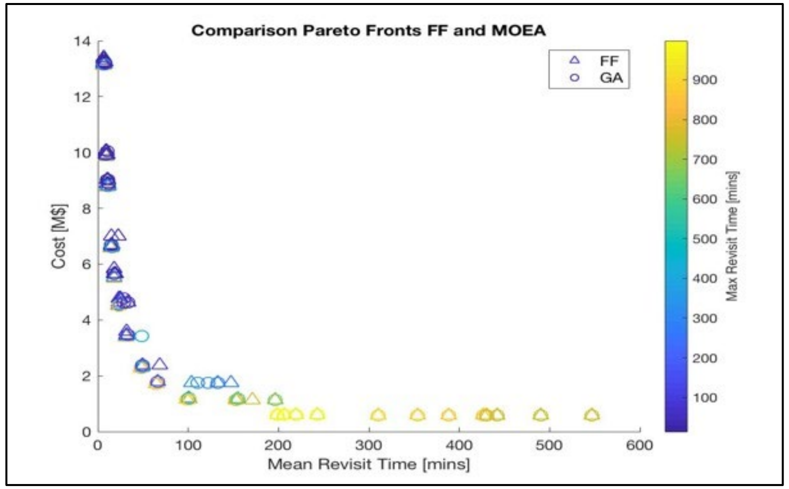
AIST-14 & -16/Grogan (Stevens) –
Trade-space Analysis Tool for Constellations (TAT-C)

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



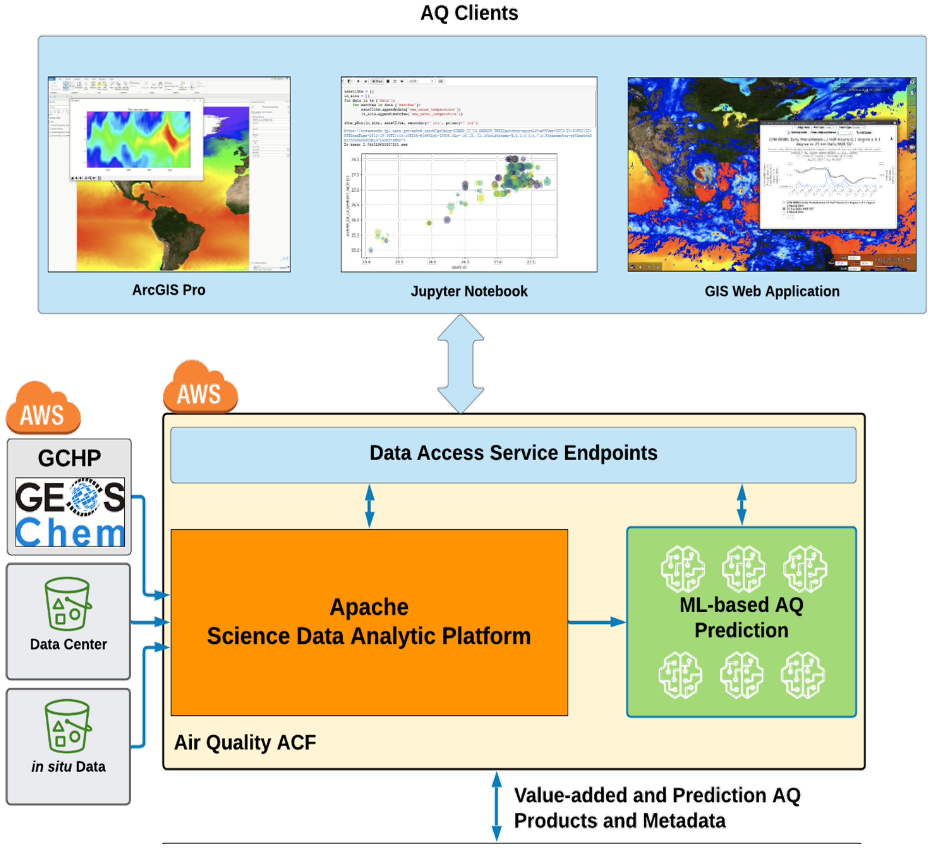
Evolutionary algorithm enables searching much larger and richer design spaces with hybrid architectures combining satellites at various altitudes and inclinations.

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

QRS-20/Huang (JPL, City of La, WUSTL, U. CO, GMU) – *AQACF: Air Quality Analytic Collaborative Framework*

This project develops an Analytic Collaborative Framework (ACF) for Air Quality that can perform analyses and projections of air quality across multiple synergistic data sets and models. It will demonstrate an analysis application focusing on air pollution in large cities (e.g., Los Angeles).

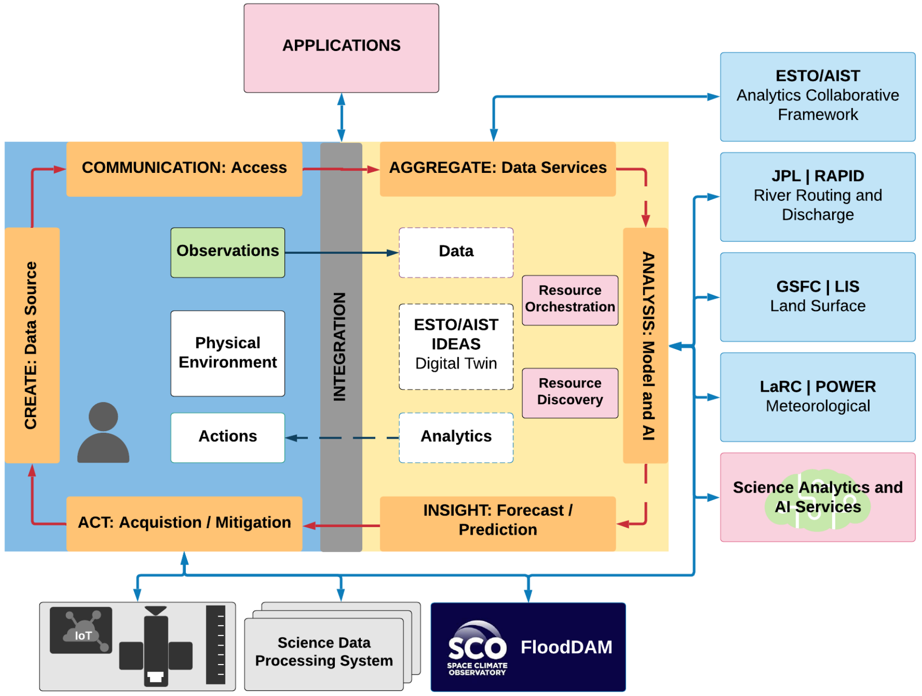


AQACF platform integrates data, analytics, and models for air quality prediction and analysis.

- Build on the Apache Science Data Analytics Platform (SDAP) architecture to harmonize air quality data sets and analytics.
- Formalize inputs to prediction models, GEOS-Chem, and GCH, as well as models to be developed through this AQACF effort
- Integrate with cloud-provisioned ML services
- Demonstrate AQACF data access and analytic services through integration with Jupyter Notebooks
- Provide a built-in toolbox of algorithms and workflows for on-demand analytics
- Develop interfaces to enable predictive models of air quality to be driven by harmonized data from multiple instruments
- Develop web-based application for data access, interactive visualization and analytics

QRS-21/Huang (NASA JPL, GSFC, LaRC with CNES) – *ESTO/AIST Integrated Digital Earth Analysis System (IDEAS)*

***IDEAS is a NASA ESTO/AIST Earth System Digital Twin** project that bridges the physical environment and its virtual representation by continuously assimilating new observations to improve forecast and prediction for integrated science and decision support.*



IDEAS – Digital Twin for Water Cycle and Flood Detection and Monitoring

Jet Propulsion Laboratory California Institute of Technology · Goddard Space Flight Center · Langley Research Center

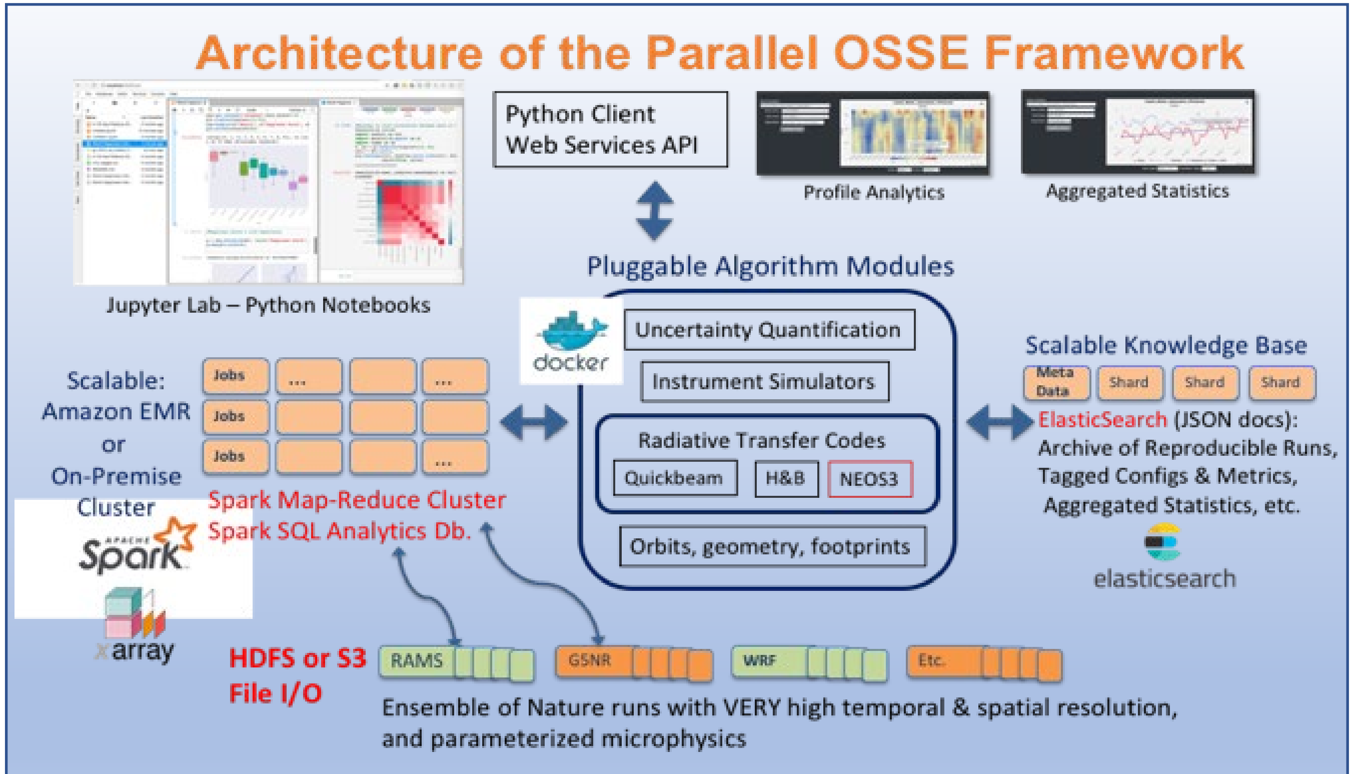
- Using water cycle and flood analysis as the prototype application to integrate NASA, CNES, and Space Climate Observatory (SCO) data and science
- Multi-Agency and Multi-Center partnership
- Advanced numerical models and analysis
  - **JPL's RAPID**: Routing Application for Parallel computation of Discharge
  - **GSFC's LIS**: Land Information System
  - **LaRC's POWER**: Prediction of Worldwide Energy Resources
  - **CNES and SCO's FloodDAM:** Automated service to reliably detect, monitor and assess flood events globally
  - Integration with **NASA IPCC Sea Level Prediction** data for coastal flooding
  - Joint developed and trained flood detection and prediction machine learning algorithms
  - Promote and advance interoperable standards
- Improve the Machine Learning flood prediction model (Huang *et al.* 2020) from the JPL-CNES Joint Data Science pilot
- Scenario-based prediction for infrastructure and population impacts
- Ongoing formulation and planning with CNES and SCO's FloodDAM
- Next interchange will be on October 6, 2021 and project is expected to start soon after

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/Moisan (GSFC) and Zhang (SMU) –
Using AI for Facilitating the Development of Novel Science Data Processing Workflows and Algorithms

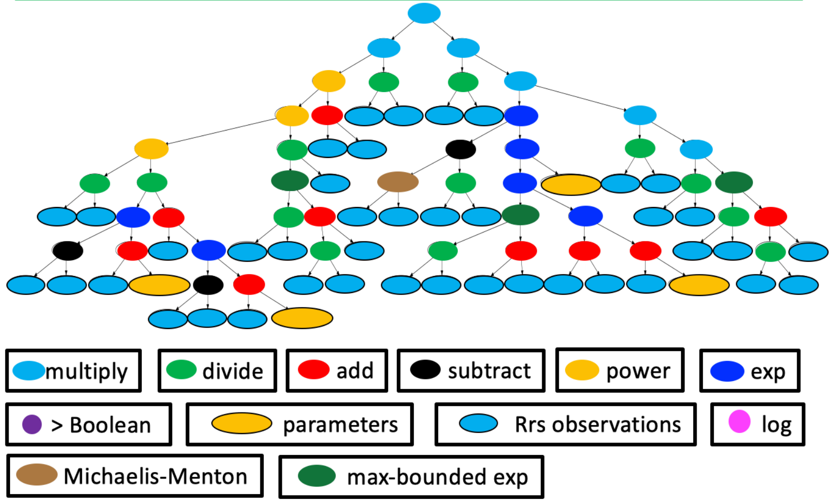
Moisan (GSFC)/"NASA Evolutionary Programming Analytic Center (NEPAC)"

Develops 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)

A GPCODE-generated Chl-a Algorithm (10 levels)

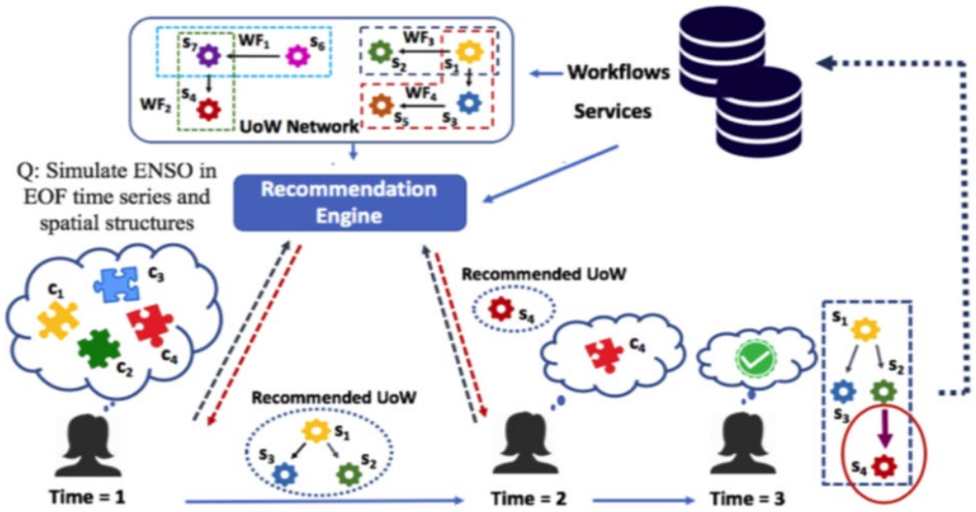


Improved Ocean Color Algorithms and Climate Data Records Using Machine Learning/Genetic Programming

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

Builds 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

ESTO
Earth Science Technology Office