AIST Overview

Advanced Information System Technology Fueling Innovation for Earth Observing

A $14M/year investment in information technology that NASA Earth Science will need in the 5-20 year timeframe

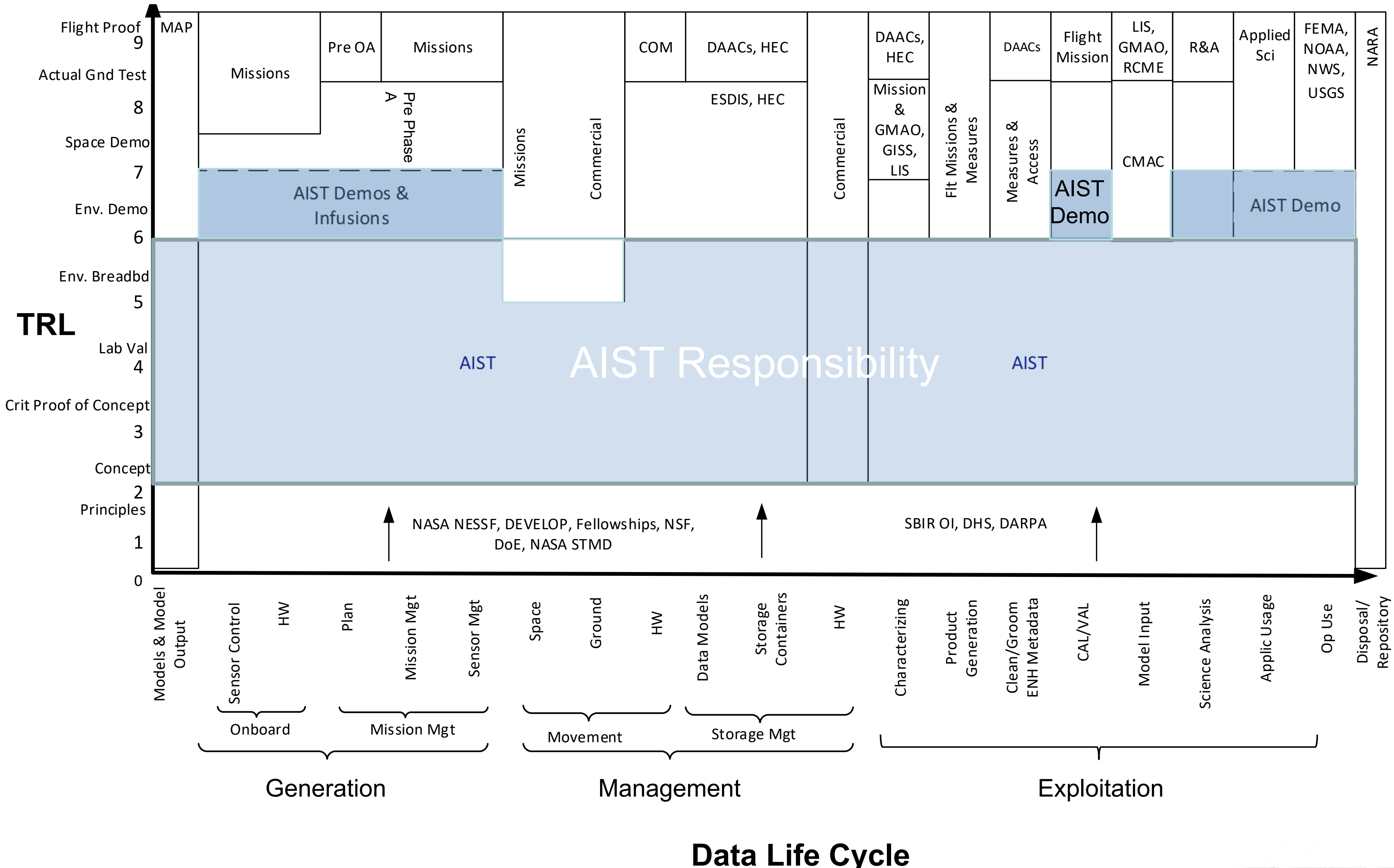
Mike Little
AIST Program Manager
November, 2018

Outline

- Program Goals, FY18 objectives and strategy
- Technology Thrusts
- AIST-18 Strategy
- Technology Infusion

How to View AIST

TRL v. Data Life Cycle



AIST Objectives

- AIST Goals
  - Reduce the risk, cost, size, and development time of Earth Science Division (ESD) space-based and ground-based information systems;
  - Increase the accessibility and utility of science data; and
  - Enable new observation measurements and information products.
- AIST 18 Objectives
  - Solicit Competed Research
    - Understand Decadal Survey guidance
    - Acquire Center/HQ input and partnership
    - Develop acquisition strategy and limit scope to key focus areas
    - Draft and approve for release
  - Improve partnerships with NASA Research and Applied Sciences to develop useful advanced info technology
    - Use of modern commercial technology: cloud computing, GIS, analytic tools
    - Focus advanced analytic techniques on supporting NASA ESD R&A and Applied Science
    - Assess the needs for multi-vantage point Sensor Web-based Observing Strategy

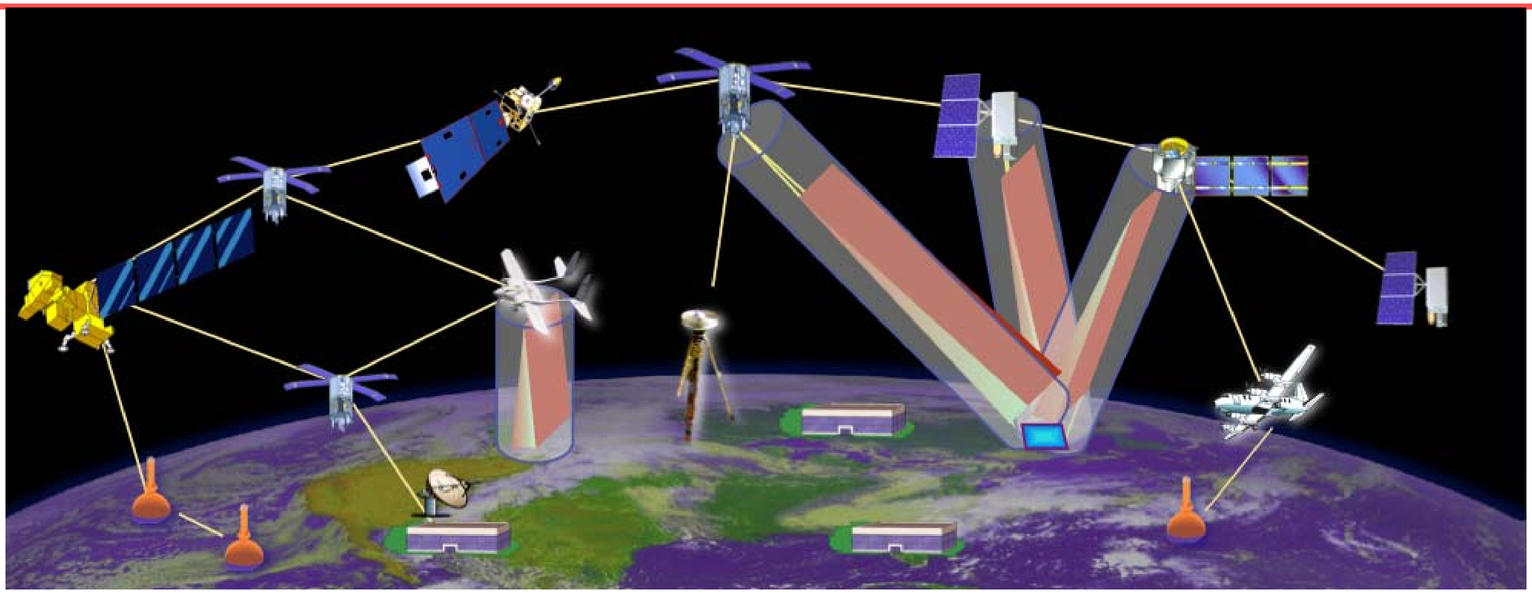
FY18 Primary Technology Thrusts

- Sensor Web Observing Strategy
  - Constellations of sensors from different vantage points
  - Designing a complete architecture to create a unified picture of a phenomenon
    - Forecast models as a measure of quality of understanding
    - Non-NASA sources of data or relevant services
- Analytic Center Framework to determine needs for tool integration
  - Focus on supporting Science Investigations
    - Allow maximum discretion on part of Science PI in using data and tools
    - Reduce the repetitive work in data access and integration
  - What tools are needed and how to accelerate development through AIST18
- Related Science support questions
  - How do we make an objective and quantitative comparison of multi-dimensional data?
  - How do we measure science value?

Program Strategy for AIST

- Solicit competed research
  - Public competition with peer review and programmatic alignment
- Organize Projects around technology thrusts
  - Technologies needed for Sensor Web Observing Strategy
  - Tools and techniques fitting into Analytic Center framework
- Optimize technology infusions identified in proposals
  - Increase collaboration with R&A and AS Program Managers
  - Solicit R&A and Applied Sciences partners for input
  - If subsequent opportunities arise, augment Awards to support new technology infusion opportunities
- Usability of commercial services to support Research and Applied Sciences
  - US and non-US

Sensor Web Observing System



- Technology advances have created an opportunity to make new measurements and to continue others less costly
  - Smallsats equipped with science-quality instruments
  - Machine Learning techniques permit handling large volumes of data
- New Observing Strategies
  - Sensor webs producing data integrated from multiple vantage points
  - A unified picture of the physical process or natural phenomenon

Earth Science Use Cases

- Sensor webs for improving model prediction
  - Integrate models with in situ, airborne and orbital instruments
  - Operational and research models
- Real-time targeting of transient and transitional phenomena
  - In situ triggering of observing system
  - Control steerable instruments
  - Train configuration prolongs observations of an event
  - Viewing an event from multiple angles
- Integrated sensor webs for phased arrays
  - Beat down error with statistics
  - Improve resolution with multi-node instruments

SWOS Candidate Science Customers

- Hydrology
  - River flow and Flooding
  - Snow fall in 3D
  - Aquifer degradation
- Precipitation
  - Extreme precipitation events
- Cryosphere
  - Glaciers changes
  - Sea Ice changes
- Urban Air Quality Events
  - At block-level resolution (vertical and horizontal)
- Biodiversity
  - Migrations
  - Invasive species
  - Transient diurnal phenomena
- Solid Earth and Interior
  - Landslides
  - Plate movement
  - Volcanic activity
  - Interior magma movement

ESTO Areas of Interest

- Integrated Observing System
  - Orbital Sensors & Platforms
    - OSP Onboard Processing
      - OSPOb Hardware
      - OSPOb Software
      - OSPOb Comm
    - Orbital Sensors
    - Orbital Platform
  - Airborne Sensors & Platforms
    - ASP Onboard Processing
      - ASPOb Hardware
      - ASPOb Software
      - ASPOb Comm
    - Airborne Sensors
    - Airborne Platform
  - In Situ Sensors & Platforms
    - ISP Onboard Processing
      - ISPOb Hardware
      - ISPOb Software
      - ISPOb Comm
    - In Situ Sensors
    - In Situ Platform
  - Science Data Subsystem
    - SD Processing
      - SD Intercal
      - SD Archive
    - SD Analytic Center
  - Mission/Campaign Operations Function
    - Command & Control Subsystem
    - Comm Subsystem
    - Mission/ Campaign Ops Software

ESTO Areas of Interest

AIST Projects – Sensor Web Observing System

PI Name	Org	ESTO ID #	Project Title	Start Date	Function
Jacqueline LeMoigne-Stewart	GSFC	16-107	TAT-C ML	8/16/17 8/15/19	Mission Planning
Joel Johnson	OSU	16-60	A multi-platform mission planning and operations simulation environment for adaptive remote sensors	8/16/17 8/15/19	Planning/ OSSE
Barton Forman	UMD	16-24	A mission planning tool for next generation remote sensing of snow	9/1/17 8/31/19	OSSE
Matthew French	USC/ ISI	16-31	SpaceCubeX: On-board Processing for Distributed Measurement and Multi-Satellite Missions	9/1/17 8/31/19	Onboard Processing
Ved Chirayath	ARC	16-*48*	NeMO-Net - The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment	9/1/17 8/31/19	Integrate Airborne
Dara Entekhabi	MIT	16-49	Autonomous Moisture Continuum Sensing Network	9/1/17 8/31/19	Integrate In Situ
Jonathan Zender	UC Irvine	16-63	JAWS: Justified AWS-like data throughput workflow enhancements that ease access and add scientific value	10/1/17 9/30/19	Integrate In Situ
Sreeja Nag	ARC	NIP	Autonomous Scheduling of Agile Spacecraft Constellations for Rapid Response Imaging	4/1/18 3/31/21	Instrument steering
Rich Doyle	JPL	QRS-16-06	High Performance Spaceflight Computing (HPSC)	9/15/16 9/15/20	Onboard Processing
Mike Lowrey	ARC		Neuromorphic Computing Applications	6/30/18	Onboard Processing
Petya Campbell	UMBC	14-90	Next Generation UAV Based Spectral Systems for Environmental Monitoring	6/1/15 11/30/18	Integrate airborne instruments
Matt McGill	GSFC	QRS-18-08	Slingshot Lidar Science Data Processing using Machine Learning	5/1/18 9/30/18	Onboard Processing
Rohit Mital	SGT	QRS-18-03	Block-chain test bed	6/1/18 9/30/18	Security
Mahta Moghaddam	USC	NESSF 18-399	Software Defined Radar Platform Development for Dynamically Configurable Multistatic and MIMO Smart Sensor Networks		Integrate In Situ instruments
Matt McGill	GSFC	QRS-18-08	Slingshot Lidar Science Data Processing using Machine Learning	5/1/18 9/30/18	Onboard Processing

Analytic Center as a Framework

Focus on the Science User

User
- Project Definition
- Plan for Investigation

Project Work Environment

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

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

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

Tropical Cycle Information System (TCIS)

Tropical Cyclone Research Program

- Project Definition
- Plan for Investigation

Data

- Integrated web portal
- Hurricane List & Tracks
- HS3 Field Campaign
- SMAP
- SST
- TPW
- WIND
- ECMWF forecast output
- HAMSR
- GOES-East
- GOES-West
- Microwave Rain Signature

Project Work Environment -JPL-

Tools

- Integrated through web portal
- Re-use requires TCIS Team
- NEOS3 instrument simulator
- Statistical evaluation
- Storm Structure
- Environment vertical structure
- Visualization
- Models

Computational Infrastructure

Storage

- Local Storage

Computing

- Local systems
- AWS through AMCE

POC: svetla.m.hristova-veleva@nasa.gov
https://tropicalcyclone.jpl.nasa.gov/

NASA Earth Exchange (NEX) as an Analytic Center

Land Change/Use Community

- Project Definition
- Plan for Investigation

Data

- Built into website
- Dataset Sources
  - Landsat
  - Sentinel 1A
  - Modis
  - ASTER
  - TRMM
  - AVHRR
- Climate Datasets
- Land cover
- Digital Elevation Map
- STATSGO Soils
- USDA Aerial NAIP
- And others

Project Work Environment -NEX-

Tools

- Models
  - Tops
  - Biome-BGC
  - LPJ Dynamic Global Model
- Sandbox for small scale experiments
- Analysis
  - R and python based tools
  - Matlab VIIRS HDF5 swath conversion
- Workflow: Jupyter Notebook

Computational Infrastructure

Storage

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

Computing

- NASA Advanced Supercomputing (NAS)
- Amazon Web Services (AWS) Public Cloud
- Ames Quantum Computer (D-wave)

POC: rama.nemani@nasa.gov
https://nex.nasa.gov/nex

AIST Tasks - Analytic Center

PI Name	Org	Prop #	Project Title	ESTO PM	End Date	Machine Learning Elements
Kamalika Das	UC Santa Cruz	115	Uncovering Effects of Climate Variables on Global Vegetation	Oza	5/31/17	Symbolic regression
Milton Halem	*UMBC*	*96*	Computational Technologies: Feasibility Studies of Quantum Enabled Annealing Algorithms for Estimating Terrestrial Carbon Fluxes from OCO-2 and the LIS Model	Cole	5/31/17	RBM, CNN, data assimilation
Hook Hua	JPL	109	Agile Big Data Analytics of High-Volume Geodetic Data Products for Improving Science and Hazard Response	Norton	5/31/17	Fault recognition and Science processing redirection
Thomas Huang	JPL	28	OceanXtremes: Oceanographic Data-Intensive Anomaly Detection and Analysis Portal	Norton	8/29/17	NEXUS as a data delivery tool - oceanography
Kristine Larson	U of CO	4	AMIGHO: Automated Metadata Ingest for GNSS Hydrology within OODT	Hines	5/31/17	Sensor characterization
Victor Pankratius	MIT	36	Computer-Aided Discovery of Earth Surface Deformation Phenomena	Little	5/31/17	Computer aided discovery
Chaowei Yang	GMU	82	Mining and Utilizing Dataset Relevancy from Oceanographic Dataset (MUDROD) Metadata, Usage Metrics, and User Feedback to Improve Data Discovery and Access	Cole	5/31/17	Natural language processing, CNN, SVM, deep learning
Tomasz Stepinski	Cincinnati	27	Pattern-based GIS for Understanding Content of very large Earth Science datasets	Quam	6/31/17	Classification and similarity
Jonathan Gleason	LaRC	95	Ontology-based Metadata Portal for Unified Semantics (OlyMPUS)	Oza	1/31/17	Precision ontology foundation
Constantine Lukashin	LaRC	14	NASA Information And Data System (NAIADS) for Earth Science Data Fusion and Analytics	Murray	1/31/17	Scientific enterprise service bus
Aashish Chaudhary	Kitware	65	Visualization Pipelines for big-data on the NASA Earth Exchange (NEX) Prototyping Agile Production, Analytics	Hines	3/31/17	Workflow
Martyn Clark	UCAR	88	Development of Computational Infrastructure to Support Hyper-resolution Large-ensemble Hydrology Simulations from Local-to-Continental Scales	Hines	4/30/17	Assimilation and ensembles
Kwo-Sen Kuo	Bayesics	56	DERECHoS: Data Environment for Rapid Exploration and Characterization of Organized Systems	Little	4/30/17	Foundation for data delivery
Seungwon Lee	JPL	32	Climate Model Diagnostic Analyzer	Norton	4/30/17	
Christian Mattman	JPL	34	SciSpark: Highly Interactive and Scalable Model Evaluation and Climate Metrics for Scientific Data and Analysis	Norton	4/30/17	Foundation for data delivery

AIST Tasks - Analytic Center

PI Name	Org	Prop #	Project Title	ESTO PM	End Date	Machine Learning Elements
Ved Chirayeth	ARC	AIST-QRS-16-0004	MiDAR-fused Supervised Machine Learning (SML)	Hines	1/31/17	Use of high resolution training sets to improve global scale moderate resolution data
Chris Mattman	JPL	AIST-QRS-16-0007	Deep Web Search Analytics	Norton	10/31/16	Text analytics
John Readey	HDF Group	ACCESS15 -0031	Object Store-based Data Service for Earth System Science	Hines	5/31/17	Foundation for data storage to improve access
Yehuda Bock	Scripps	AIST-QRS-16-0010	Latency test of realtime warning systems in AMCE Cloud Computing	Quam	9/30/17	Event Detection in near real time
Chris Lynnes	GSFC	AIST-QRS-16-0001	Experiment with Data Containers in ESDIS Context	Little	12/20/16	Evaluation of alternative storage models for data to enable analysis

Fostering Technology Infusion

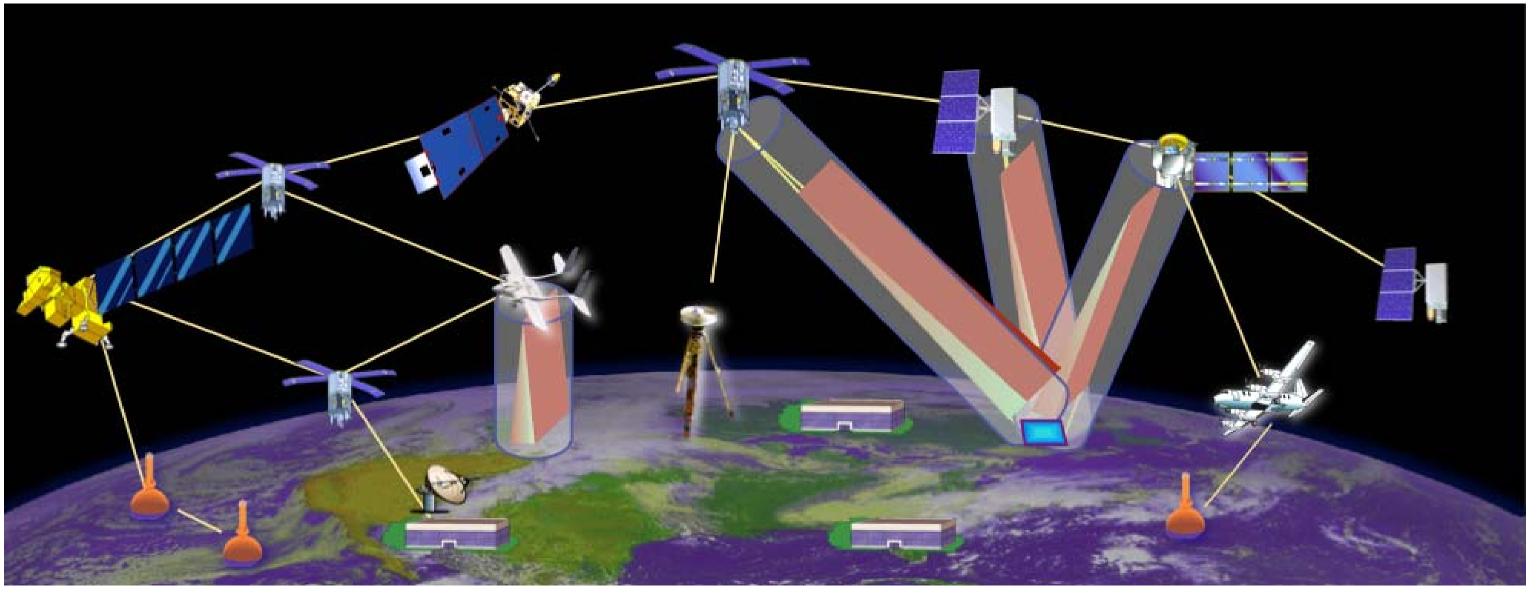
- Focus technology projects on solving science problems
  - Solicitation provisions
  - Augmentations to demonstrate proof of concept to science adopters
  - Machine Learning (Analytic Center) Workshops
  - Sensor Web (New Observing Strategy) Workshops
  - Regular conversation with HQ Program Scientists and Applied Sciences PM's
  - What science problems are out of reach?
- Increase awareness/acceptance by science community
  - Public Cloud Computing (AMCE)
  - Machine Learning tools for analyzing data
  - Geographic Information System (GIS)
  - Commercial Analysis Services (i.e., Decartes Labs)
  - Active Intercalibration to enable integrated datasets

Backup

FY18 Primary Technology Thrusts

- Sensor Web Observing Strategy
  - Constellations of sensors from different vantage points
  - Designing a complete architecture to create a unified picture of a phenomenon
    - Forecast models as a measure of quality of understanding
    - Non-NASA sources of data or relevant services
  - Command and control structures and tools
    - Scaling mission operations to the constellation level
    - Potential issues for traffic control in Low Earth Orbit
    - Realistic and useful computer security
    - High degrees of autonomy to be responsive to changing conditions
    - Reduce the current labor intensive operations
- Analytic Center as a framework to determine needs for tool integration
  - Focus on supporting Science Investigations
    - Avoid force-fitting science into a single approach
  - What tools are needed and how to accelerate development through AIST18
- Related Science support questions
  - How do we make an objective and quantitative comparison of multi-dimensional data
  - How do we measure science value?

So…What is a sensor web?



"A sensor web is a distributed system of sensing nodes that are interconnected by a communications fabric and that functions as a single, highly coordinated, virtual instrument. It autonomously detects and dynamically reacts to events, measurements, and other information from constituent sensing nodes and from external nodes (e.g., predictive models) by modifying its observing state so as to optimize science information return."

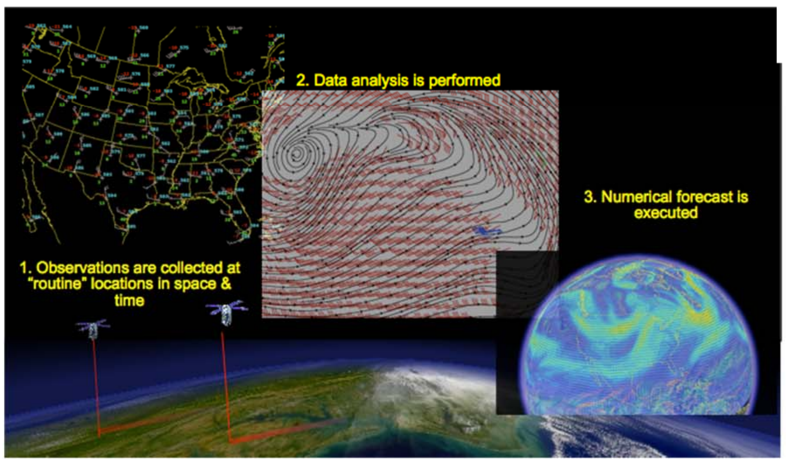
- Steve Talabac et al, 2003

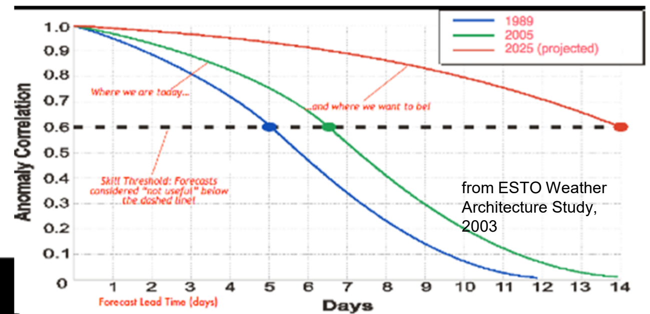
Sensor Webs for Improving

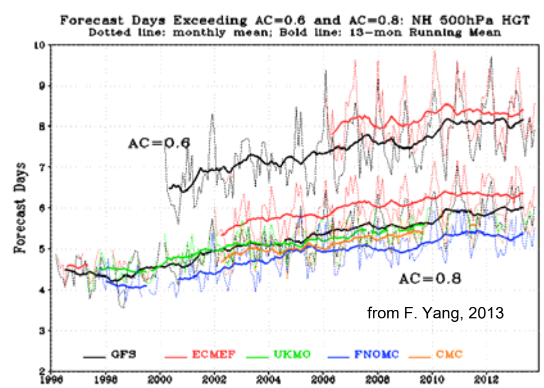
Weather Prediction?

2006: Technology project launched to determine if sensor webs could provide a "revolutionary" improvement in the skill of numerical weather forecasts

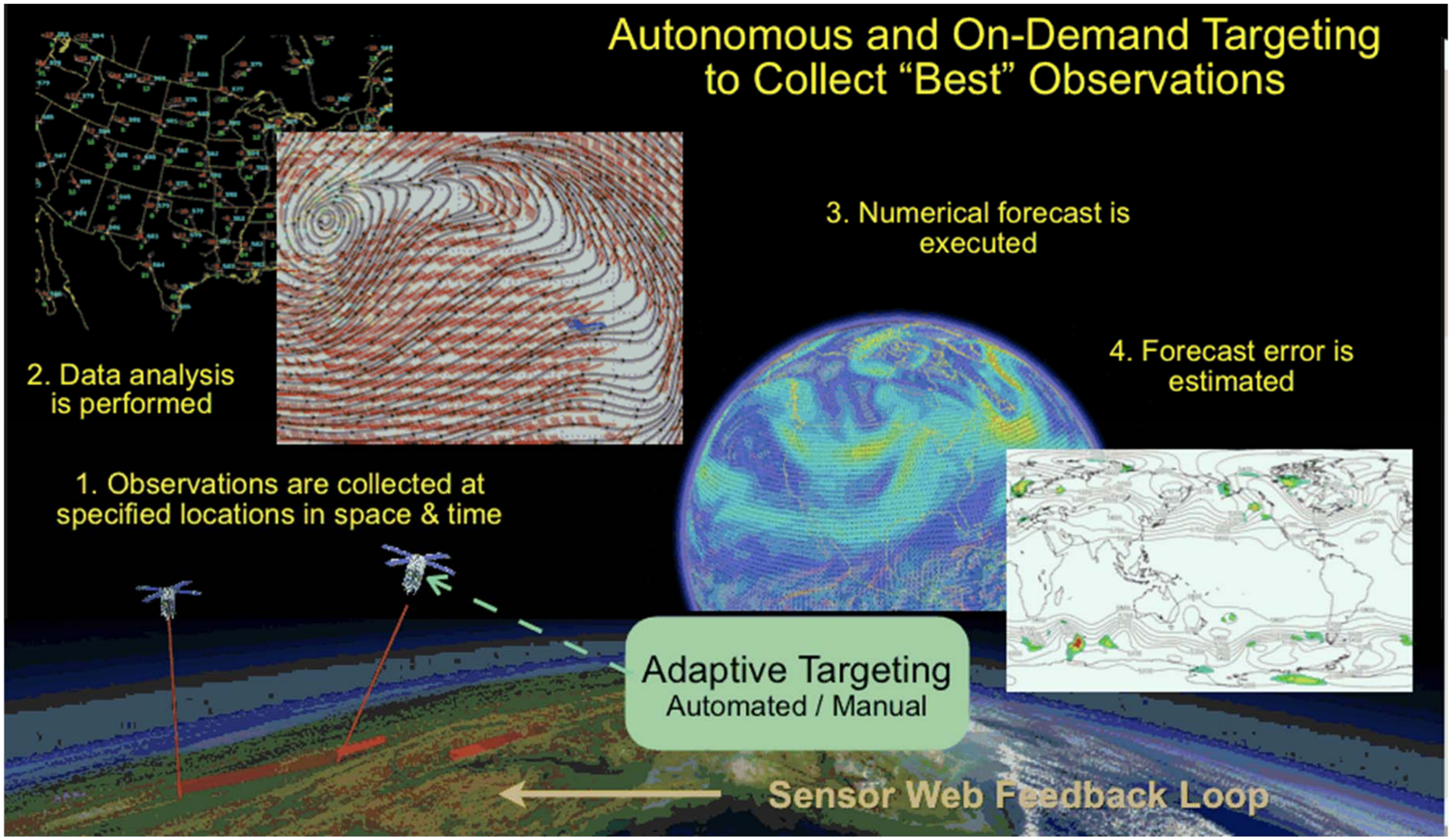
Could 7-day skill in 2005 improve to 14-day skill by 2025?







Real-Time Data for Adaptive Targeting



Potential Concept of Operations:

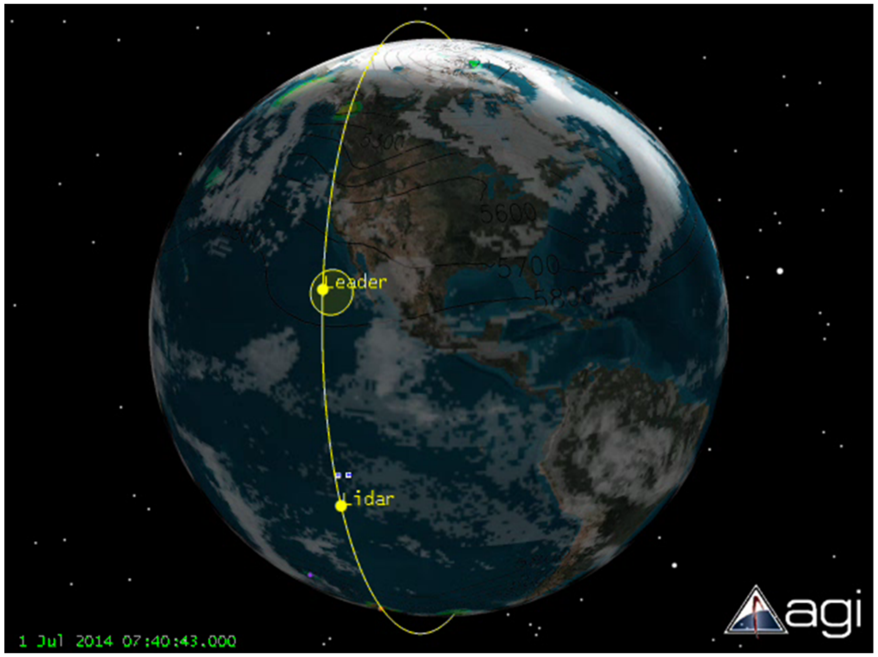
Adaptive targeting with a wind lidar

Movie depicts mission CONOPS -- wind lidar working with an operational atmospheric model's first-guess field that identifies regions that are sensitive to forecast error

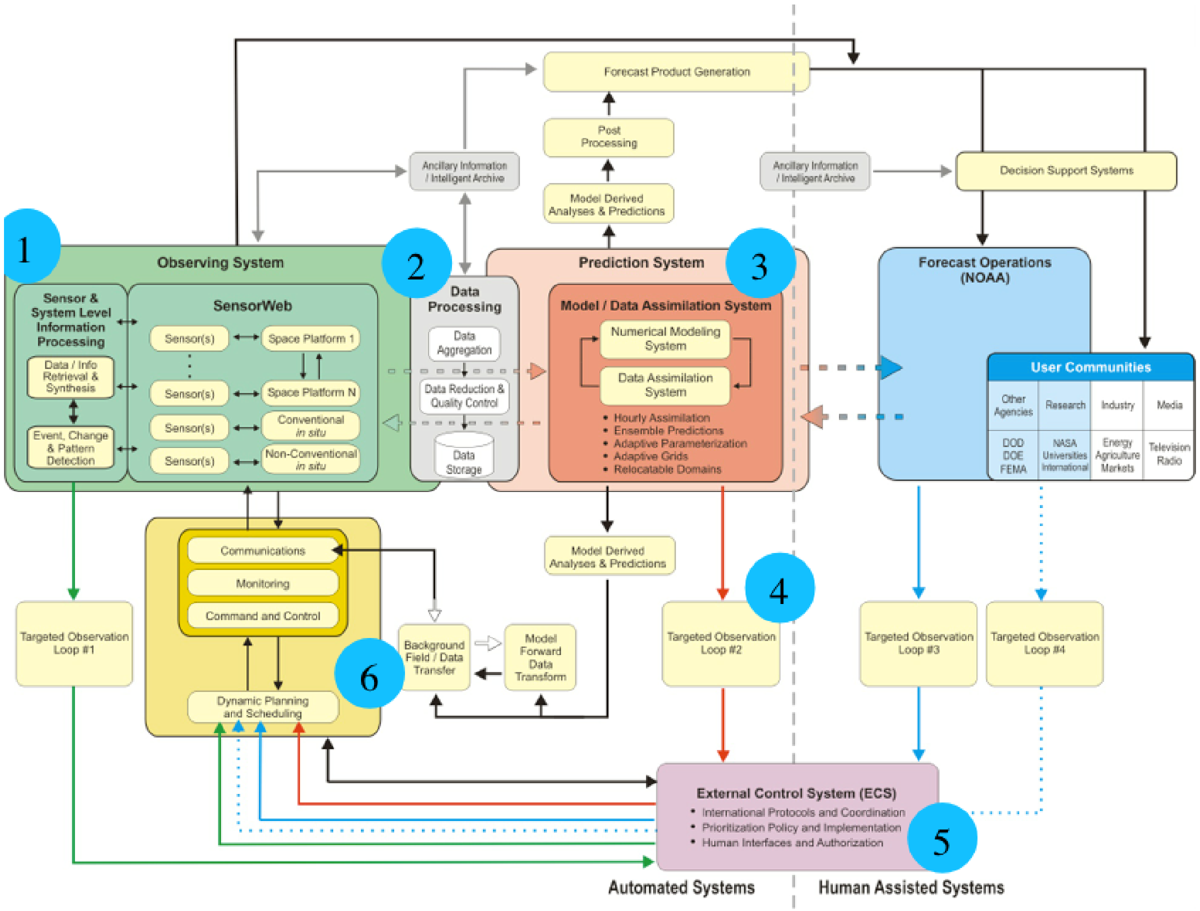
Spacecraft will slew toward sensitive regions and lidar is placed in a high data rate collection mode

"Leader" spacecraft is included to depict how optimization for cloud-free lines of sight could be performed

(Operational CONOPS would be more complex)



Proposed Sensor Web Architecture



Use cases are executed based on the proposed architecture to determine technology gaps and to estimate the feasibility of a “generic” simulator (i.e., high level of reuse)

Higgins, G., Kalb, M., Lutz, R., Mahoney, R., Mauck, R., Seablom, M., Talabac, S., 2003: “Advanced Weather Prediction technologies: Two-Way Interactive Sensor Web & Modeling System”

Analytic Center Concept

- An environment for data analysis in a Science investigation
  - Tailored to the individual study
  - Stand it up when needed, archive when complete
  - Collects publication submission materials in background (data, source code, version tracking)
- Harmonizes data, tools and computational resources to permit the research community to focus on the investigation
  - Reduce the data preparation time to something tolerable through re-use
  - Catalog of optional resources (think HomeDepot shopping or AppStore)
  - Collect relevant publications
  - Provide established training data sets of varying resolution
  - Provide effective project confidentiality, integrity and availability
  - Seamless integration of new and user-supplied components and data

Analytic Center Strategy

- Socialize the concepts with the science community
  - Does it resonate with Program Scientists and Managers
- Conduct a workshop to assess needs and technology gaps
  - If it works, do more of them
- Solicit competed research
  - Accelerate the process where appropriate
  - Leverage other solicitations (NESSF, NIP)

Analytic Center Features

- Low infrastructure cost
  - Can be stood up on short notice for an individual investigation
  - Only maintained if a reason to maintain continuous availability
- Seamless integration of new components
  - User provided tools
  - User provided data
  - Local computing environment
- Comprehensive catalog of data sources and tools
  - Clear applicability (or not)
  - Shopping for data, not searching
  - Expert system as an operator aid in selecting accepted tools
- Help in using them
  - Video examples and training
  - Expert System Support

Earth Data Analytics System (EDAS) as an Analytic Center

Climate Impact

- Project Definition
- Plan for Investigation

Data

- MERRA
- ECMWF ERA
- NOAA NCEP CFSR
- JMA JRA
- UA ERA

Project Work Environment -NCCS-

Tools

- Earth System Grid Federation Framework
- Discovery & Catalog
- Work Management – Jupyter Notebook
- Data Interfaces - OpenDAP
- Analytic Tools: Min, max, sum, diff, average, rms,anomaly, std deviation
- Visualization using time series or spaceplot plot routines

Computational Infrastructure

Storage

- Local Sparc
- SGI Data Migration Facility

Computing

- Local systems
- High End Computing (DISCOVER)
- GSFC Science Cloud

POC: laura.carriere@nasa.gov
https://www.nccs.nasa.gov/services/Analytics

Analytic Center Framework FY18

- Define fundamental or threshold characteristics
  - Inputs from Science community
  - ConOps and Common elements
- Inventory Existing Examples
  - Lessons Learned, weaknesses, corrective action
  - Inventory technologies for each tool type
  - Inventory viable data storage containers and models
- ID Technology deficiencies
  - From experiments
  - Solicit AIST18

- Socialize inside the science community
  - Workshop Examples
    - Atmospheric Composition
    - Hydrology
    - Cryo
    - Biodiversity
  - Opportunistic
    - GIS– LaRC GIS team + Christine White in ESRI
    - AWS Experiment as an AC
    - Commercial: Descartes Labs, Radiant Earth, Digital Globe, Esri
  - Talk About it per comm plan
- Form Partnerships
  - Jared Entin, Gerald Bawdin, Tom Wagner, Barry Lefer, Woody Turner/Allison Leidner

Candidate Useful ML Tools

- Implementations
  - TensorFlow
  - TensorFire (ML in the browser using jess)
  - arc-GIS server (Esri)
  - MathLab (Mathworks)
- Algorithms
  - Logistic regression
  - Symbolic regression
  - Random forest
  - Convolutional Neural Networks (CNN)
  - Deep Neural Networks (DNN)
  - Case-based reasoning
  - Restricted Boltzman Machines

AIST-16 Tasks-Machine Learning

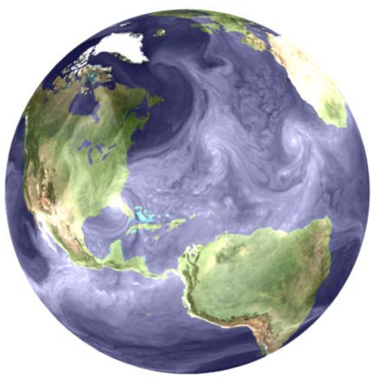
PI Name	Org	Prop #	Project Title	ESTO PM	Start Date	End Date	Machine Learning Elements
Victor Pankratius	***MIT***	***48***	Computer Aided Discovery and Algorithmic Synthesis for Spatio-Temporal Phenomena in InSAR	Oza	8/1/17	**7/30/19**	Computer Aided Discovery
Jacqueline LeMoigne-Stewart	**GSFC**	**107**	TAT-C ML	Quam	8/1/17	**7/30/19**	Design space iterator
Branko Kosovic	**NCAR**	**79**	Fuel moisture content for improved fire prediction	Hines	6/1/15	**5/31/17**	Data assimilation
Andrew Michaelis	**ARC**	**137**	Framework for Mining and Analysis of Petabyte-size Time-series on the NASA Earth Exchange (NEX)	Oza	8/1/15	**7/31/17**	Time series analysis
Barton Forman	**UMD**	**24**	A mission planning tool for next generation remote sensing of snow	Quam	9/1/17	**8/30/19**	OSSE and assimilation
Ved Chirayath	***ARC***	***48***	NeMO-Net - The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment	Oza	9/1/17	**8/31/19**	High resolution training sets improve moderate resolution imagery
Dara Entekhabi	**GSFC**	**49**	Autonomous Moisture Continuum Sensing Network	Hines	9/1/17	**8/31/19**	
Milton Halem	**UMBC**	**91**	An Assessment of Hybrid Quantum Annealing Approaches for Inferring and Assimilating Satellite Surface Flux Data into Global Land Surface Models.	Little	6/1/15	**8/31/19**	Assimilation, registration
Jeffrey Morisette	**USGS**	**124**	Advanced Phenological Information System	Hines	9/1/17	**8/31/19**	
Walter Jetz	**Yale**	**92**	Software workflows for remote sensing-based biodiversity change monitoring	Cole	9/15/17	**9/14/19**	
Jonathon Hobbs	**JPL**	**30**	Simulation-based Uncertainty Quantification	Norton	10/1/17	**9/30/19**	Statistics
Martyn Clark	**NCAR**	**81**	Climate risks in the water sector: Advancing the readiness of emerging technologies in climate downscaling and hydrologic modeling	Hines	10/1/17	**9/30/19**	

Simulating Observations

from Future Observing Systems

- An **Observing System Simulation Experiment (OSSE)** is an experiment designed to assess the potential impact of planned missions on Numerical Weather Prediction. OSSEs, now widely used, were pioneered at NASA by Dr. Robert Atlas.
- OSSEs help quantify the potential benefits of an observing system before it is designed, built and launched into orbit.
- Trade-offs in instrument or orbital configurations and methods of assimilating a new type of observing system can be determined by an OSSE and ultimately result in both time and cost savings.

- A **Nature Run (NR)** is a high resolution long integration from a state-of-the-art numerical weather prediction model.
- It acts as a proxy atmosphere for OSSE's from which synthetic observations from existing and future observing systems are derived.
- It is also used as the verification or truth data set when evaluating assimilations and forecasts which use the synthetic observations.
- Available Nature Runs include: fvGCM, T511 and WRF-ARW.



Simulated Clouds from the NASA/DAO fvGCM Nature Run

OSSE Assessment

- Recent and current projects
  - Weather (GMAO/McCarty, Bob Atlas)
  - Hydrology (GSFC/Peters-Lidard)
  - Snow (UMd/Forman)
  - General purpose (Johnson/OSU)
- Need for improved OSSE technologies
  - Reduce cost and turnaround time
  - Feed into optimizing model-based engineering tools
  - Ensure simulation includes all data sources
- Approach for Improving effectiveness of OSSEs

Comparing multi-dimensional datasets

- Motivation
  - How do you validate models with observations?
  - How do you fuse data?
- Technologies needed to compare multi-dimensional datasets
  - Needs a better definition of the problem in order to develop the technologies
  - What are the requirements for integrating or comparing
- Introduced discussion at Theory of Data Systems workshop at CalTech
  - Sponsored by Statistical and Applied Mathematics Sciences Institute (SAMSI)
- FY17 and FY18 studies to define the problem enough to compete research in AIST-18