National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

An Earth System Digital Twin for Flood Prediction and Analysis

Thomas Huang, Cedric David, Jason Kang, Kevin Marlis, Miles Milosevich (intern), , Stepheny K. Perez, Wai (William) Phyo, Joe T. Roberts, Catalina Oaida Taglialetela, Mark Wronkiewicz
NASA Jet Propulsion Laboratory, California Institute of Technology

Sujay V. Kumar, Nishan Biswas
NASA Goddard Space Flight Center

Paul W. Stackhouse, David Borges, Jason Barnett, Madison P. Broddle, Bradley Macpherson
NASA Langley Research Center

Simon Baillarin, Frederic Bretar, Gwendoline Blanchet, Peter Kettig, Raquel Rodriguez Suquet
Centre National d'Etudes Spatiales

Sophie Ricci, Thanh-Huy Nguyen, Andrea Piacentini
CERFACS European Centre for Research and Advanced Training in Scientific Computation

Alice Froidevaux, Thanh-Long Huynh, Romane Raynal
QuantCube Technology

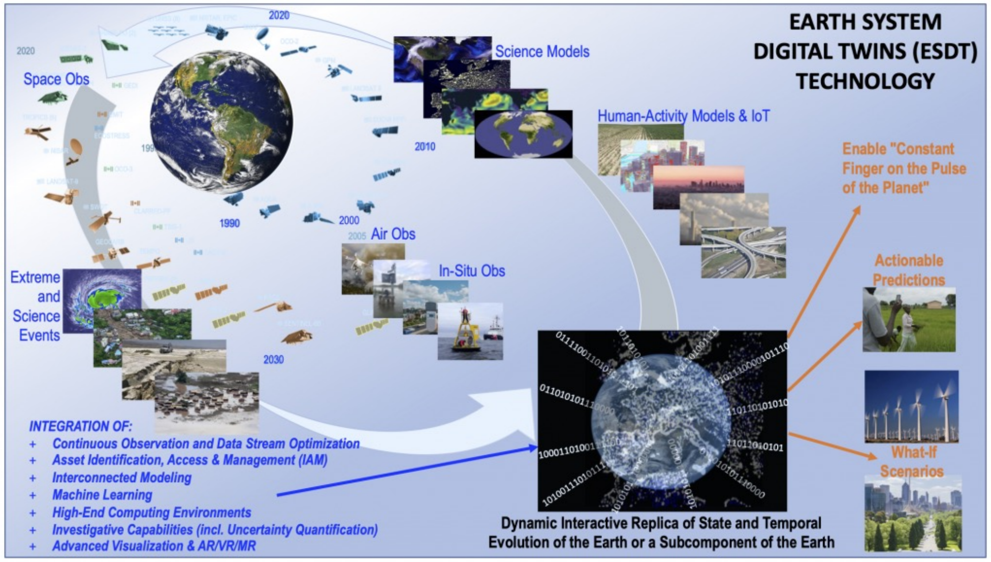
Guillaume Valladeau, Jean-Christophe Poisson
vorteX.io

Clearance: CL#22-3381

National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Earth System Digital Twins

- An **Earth System Digital Twin (ESDT)** – an interactive and integrated multidomain, multiscale, digital replica of the state and temporal evolution of Earth systems
- It dynamically integrates
  - Relevant Earth system models and simulations
  - Other relevant models (e.g., related to 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
- 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



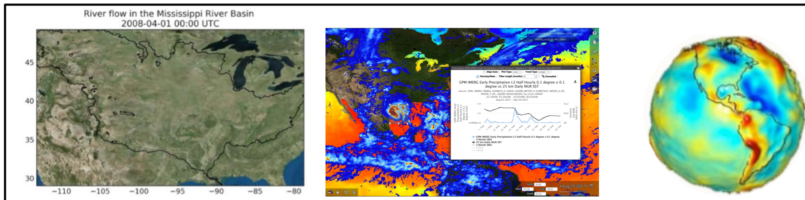
An integrated information system that, for example, enables continuous assessment of impact from naturally occurring and/or human activities or physical and natural environments

Source: https://esto.nasa.gov/aist/

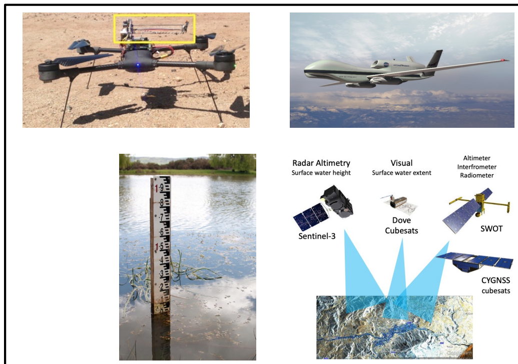
The Vision: Toward Earth System Digital Twin for Water Cycle and Flood Analysis

Technology to Bridge the Physical and Digital Environments

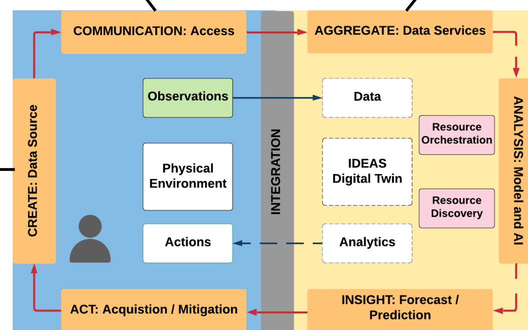
Automate Access to Many Repositories and Services

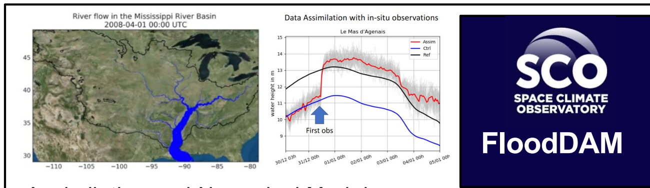


Harmonize Observation and Model Data Access and Analysis



Acquire Observation and Analysis

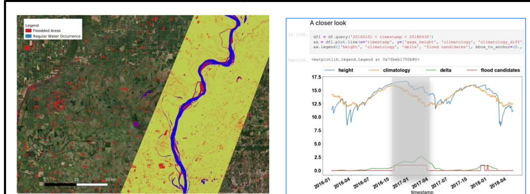




Assimilation and Numerical Models

AI-based Analysis

Decision Support and Science Planning



Forecast and Prediction

- Facilitate access, integration, and understanding of disparate datasets
- Streamline data assimilation for models and analysis
- Enable dynamic integration of new observation and analysis
- Establish interoperable ML models and data services

National Aeronautics and Space Administration

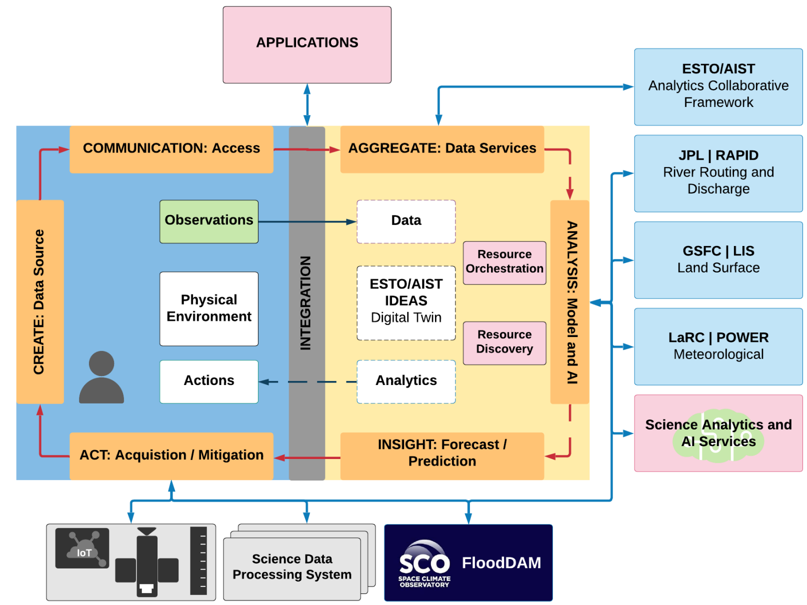
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

ESTO/AIST Integrated Digital Earth Analysis System (IDEAS)

Using water cycle and flood analysis as the prototype application

ESTO Earth Science Technology Office

- **Funded by NASA's Earth Science Technology Office (ESTO)'s Advanced Information Systems Technology (AIST) Program**
- **IDEAS -** Earth System Digital Twins project bridges the physical environment and its virtual representation - Continuously assimilating new observations to improve forecast and prediction for integrated science and decision support
- Multi-Agency and Multi-Center partnership (NASA (JPL, GSFC, LaRC) and CNES)
- **Objectives**
  - Develop a candidate software architecture for an Earth System Digital Twin (ESDT) that can coordinate services, models and observations (data) from multiple sources to analyze interacting Earth systems.
  - Prototype an application of the architecture to demonstrate three key ESDT capabilities in the context of floods and their impacts:
    - Harmonize observations and model outputs to **analyze and explore the current state of the Earth system (flooding)**
    - Coordinate models and observations to **perform predictions and what-if projections** of floods and their impacts.
    - **Federate with other ESDTs** to allow more comprehensive analyses by leveraging their data sets, models, and analytics.



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

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Vision

Bridges the physical environment and its virtual representation by continuously onboarding new observations to improve forecast and prediction for integrated science and decision support

Pittsburgh, Peter Miller

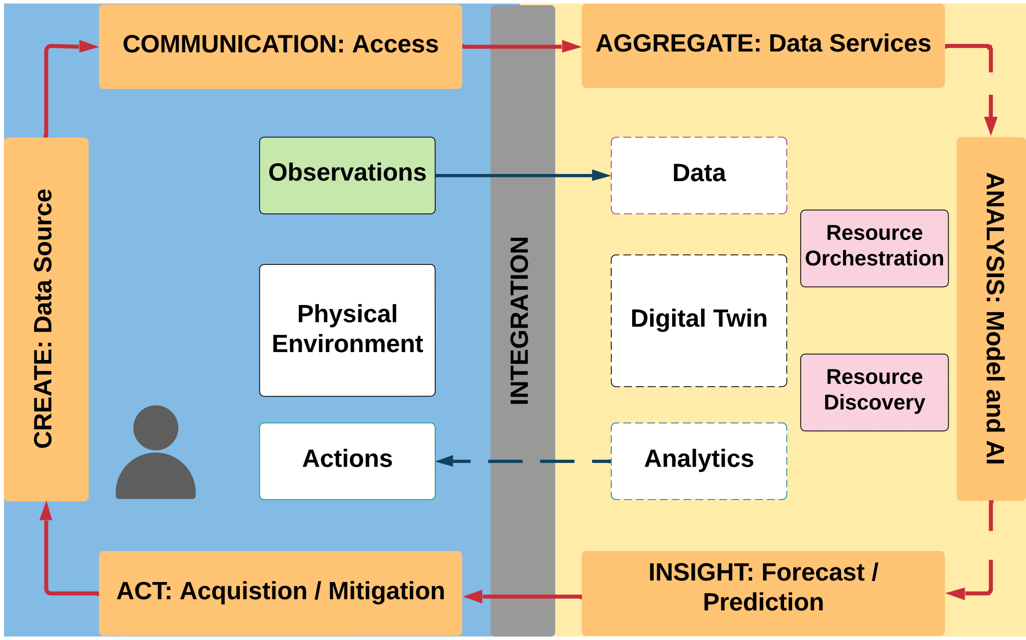
- Core building blocks
  - Resource Interface
  - Resource Orchestration
  - Resource Discovery
  - Analytics Collaborative Framework
  - New Observing Strategies
  - Multiphysics, Multiscale, and Probabilistic Models
  - Machine Learning and Deep Learning
- An Earth System Digital Twin is NOT
  - **Digital Model**: a digital version of a pre-existing or planned physical object, to correctly define a digital model there is to be no automatic data exchange between the physical model and digital model
  - **Digital Shadow**: a digital representation of an object that has one-way flow between the physical and digital object.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Earth System Digital Twin: Key Components

- **Data and Services Assets**: Supports Extract, Transform, and Load (ETL) workflow for metadata harvesting, error detection and correction, re-gridding/reprojecting, Analysis Ready Data (ARD) transformation
- **New Observation and Analysis**: Smarter method to automate onboarding relevant data
- **Integrated Multiphysics, Multi-scale, Probabilistic Models**: Automates inclusion of the latest measurements and supports scenario-based model execution
- **AI and Advanced Analytics**: enables scenario-based predictions, dynamic data acquisition, long-term prediction, data classification, process orchestration and management, etc.

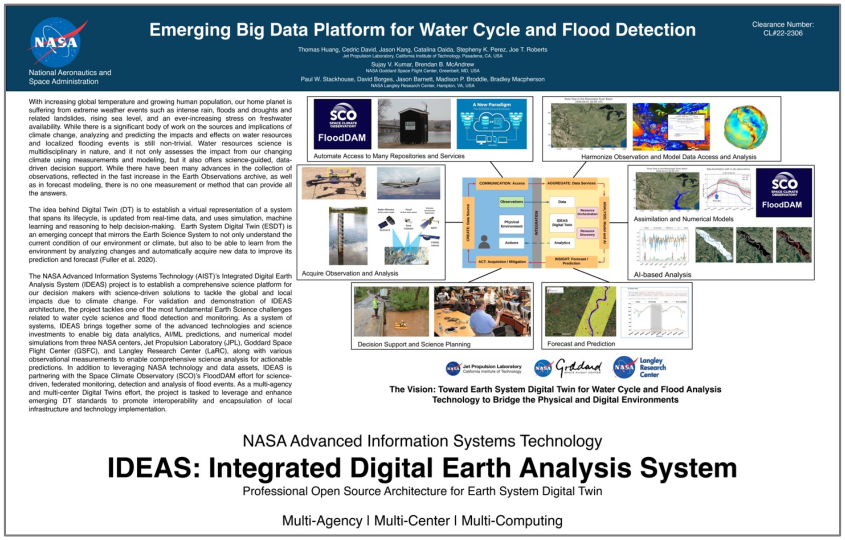


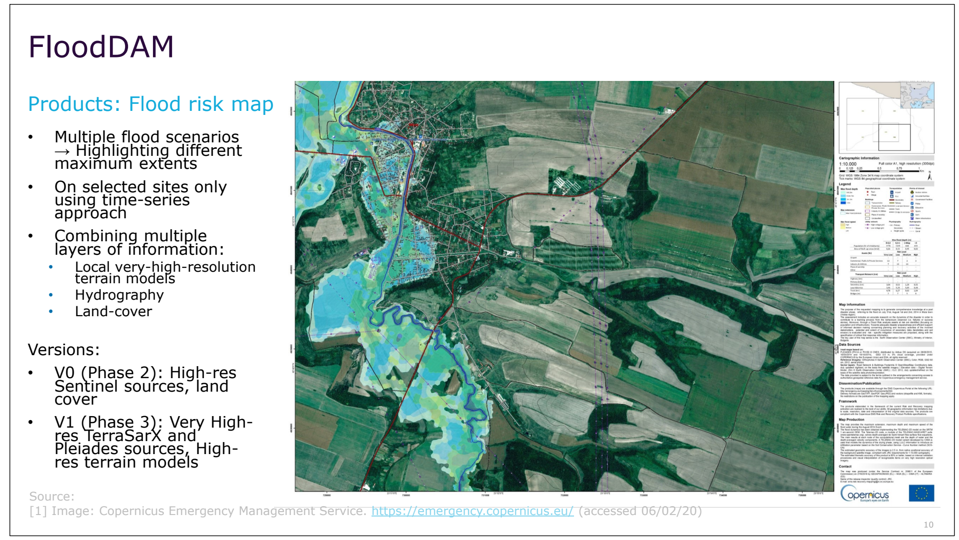
Bridges the Physical Environment and its Virtual Representation

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Federated Digital Twins – NASA IDEAS and SCO FloodDAM-DT





ESA Living Planet Symposium 2022, Bonn, Germany

FloodDAM, Space for Climate Observatory (SCO)

IDEAS and FloodDAM-DT Teams meeting, Toulouse, CNES

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Use Case: Mississippi River

Use Case	2019 Mississippi River Flooding
Period	October 1, 2018 through Aug 31, 2019 ● 2018-10-01 : 2019-02-28 - antecedent conditions ● 2019-03-01 : 2019-06-30 - flooding period ● 2019-07-01 : 2019-08-31 - beginning of recovery
Region of interest (ROI)	Mississippi River Basin (South Dakota, Nebraska, Iowa, Missouri, Tennessee, Louisiana, Mississippi), N. America -117 : -80 deg E || 27.5 : 54 deg N
Datasets of Interest (for Applications / Analysis)	● POWER downward SW and LW radiative fluxes, precipitation (rain, snow), temperature (melt timing), soil moisture ● IMERG precipitation ● SMAP soil moisture (saturation) ● MODIS snow/ice cover (timing of thaw both in south and north) ● Sentinel-1 SAR (river ice detection) ● Landsat-8, Sentinel-2 A/B imagery (river ice and flood detection) ● RAPID discharge (current + forecast) ● HLS (surface water extent/flooding) ● *SWOT river height, surface water extent ● *OPERA dynamic surface water extent ● USGS in-situ stream gauge observations ● SEDAC socioeconomic data (crop, population, infrastructure impacts)
Primary Causes	● Historic, record breaking and above normal precipitation during winter, spring, and into summer ● Compounded by river ice and snow melt with warming temperatures

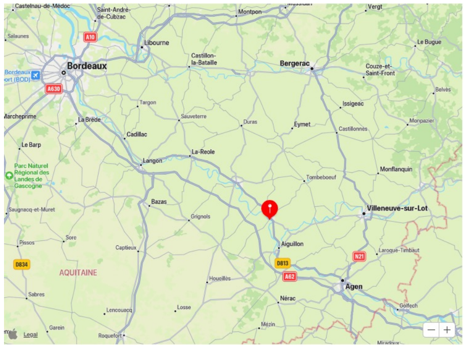
Image Credit: The New York Times

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Use Case: Garonne River

Use Case	Garonne River, France
Period	2019-12-08 : 2020-01-02 2021-01-16 : 2021-02-10 2021-12-31 : 2022-01-31
Region of Interest	Garron River: Upstream location at Tonneins : X = 437850.0 Y = 234106.6 (EPSG:27563 Lambert Sud France), lon = 0.301146886, lat = 44.38892824 Downstream location at La Réole : X = 412138 Y = 255882 (EPSG:27563 Lambert Sud France), lon = -0.034612117, lat = 44.57931175 Bbox: -0.034 : 0.30 || 44.39 : 44.58
Datasets of Interest	• POWER surface downward radiative fluxes, surface meteorological parameters (as inputs to LIS); precipitation (rain, snow), soil moisture (for analysis) • IMERG precipitation • SMAP soil moisture (saturation) • Landsat-8, Sentinel-2 A/B imagery (flood detection) • LIS surface and subsurface runoff (as input to RAPID) • RAPID discharge (current + forecast) (as input to CERFACS) • CERFACS Derived Flood Map • Sentinel-1 SAR (change in water extent) • VORTEX.IO • HLS (surface water extent/flooding) • *SWOT river height, surface water extent • *OPERA dynamic surface water extent • CERFACS Derived Flood Map
Primary Causes	Heavy winter rains



Upstream location at Tonneins
X = 437850.0 Y = 234106.6 (EPSG:27563 Lambert Sud France)
Lon = 0.301146886, Lat = 44.38892824

Downstream location at La Réole
X = 412138 Y = 255882 (EPSG:27563 Lambert Sud France)
Lon = -0.034612117, Lat = 44.57931175

Image Credit: https://earth-chronicles.com/natural-catastrophe/flooding-in-southwest-france.html

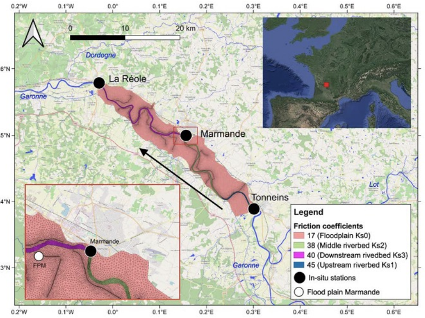
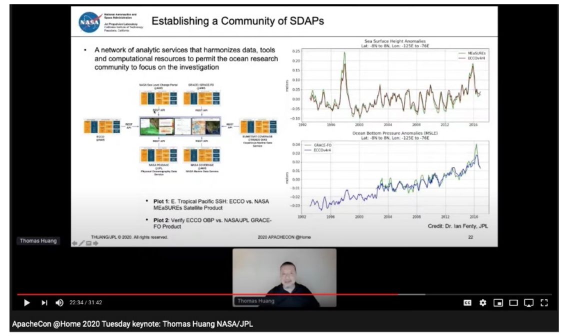
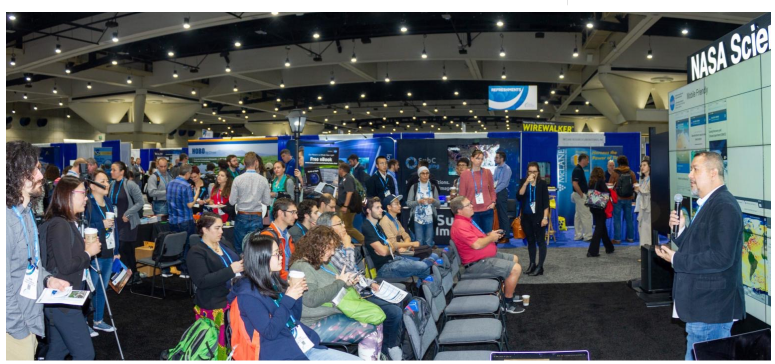


Image Credit: European Center for Research and Advanced Training in Scientific Computing (www.cerfacs.fr)

International and Community Engagement

- Engage international communities
  - Committee on Earth Observation Satellites (CEOS)
  - Group on Earth Observation (GEO)
  - United Nations Office for Disaster Risk Reduction (UNDRR)
  - United Nations Committed of Experts on Global Geospatial Information Management (UN-GGIM)
  - Open Geospatial Consortium (OGC)
- Partnership with Apache Software Foundation
- Open-Source Science
  - Share recipes and lessons learned
  - Community validation
  - Technology demonstrations
  - Inclusive and Diverse Project Management Committee (PMC)
  - Host webinars, hands-on cloud analytics workshops and hackathons





In Summary

- NASA IDEAS and SCO FloodDAM-DT
  - Development Calendar Year 2023
  - Opportunity for NASA and CNES to define and formalize the Earth Science Digital Twin architecture
  - Federated RESTful architecture: IDEAS and FloodDAM-DT can leverage each other's data, analysis, and services.
- NASA IDEAS
  - An architecture with a vision to integration existing data, analysis, science assets
  - It will be open source for the world to improve it, because open source should not be a destination, it should be in place from the beginning
  - Embrace metadata and file standards
  - The IDEAS architecture must be
    - Sustainable
    - Affordable
    - Portable
    - Extensible – from generalization to specialization
    - Federated – from global to local
  - It needs to be extensible to include other science disciplines

Agility & Resiliency

Thomas Huang

thomas.huang@jpl.nasa.gov

NASA Jet Propulsion Laboratory

California Institute of Technology

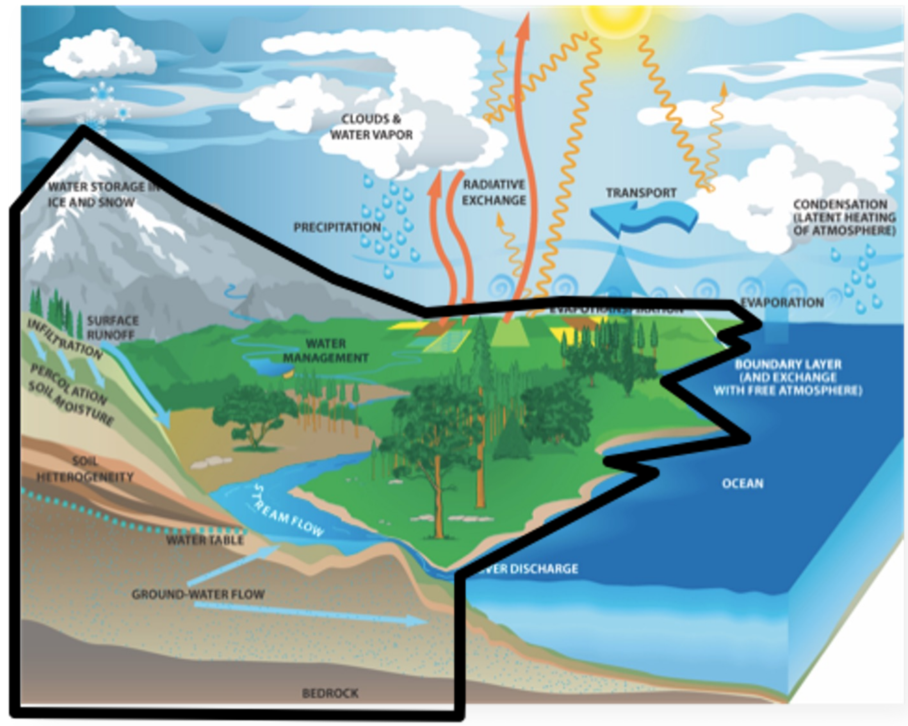
BACKUP

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Land Information System (LIS)

- A system to study land surface processes and land-atmosphere interactions
- "Use best available observations" to force and constrain the models
- Applications: Weather and climate model initialization, water resources management, natural hazards management
- Need a system viable at different spatial and temporal scales
- Be able to demonstrate the impact of observations at the scale of observations themselves
- Explicit characterization of the land surface at the same spatial scales as that of cloud and precipitation processes helps in improving the characterization of land-atmosphere interactions
- Need scalable, high performance computing support to deal with computational challenges
- Need advanced land surface models and modeling tools (data assimilation, optimization, uncertainty modeling)

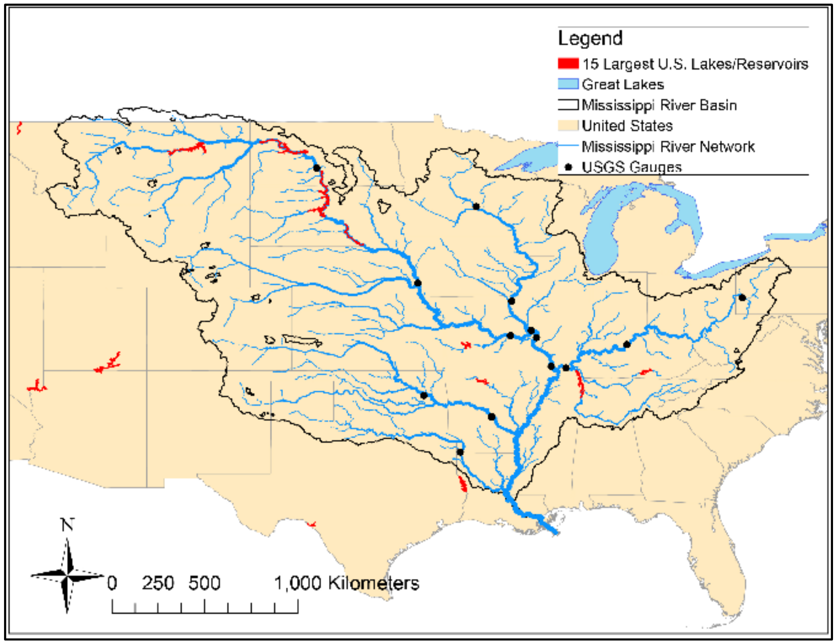


National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

RAPID – River Model

- **River Model:** Routing Application for Parallel computation of Discharge (RAPID)
  - Written in Fortran90, leverages the PETSc library (high performance computing with MPI)
  - https://github.com/c-h-david/rapid/
- **Pre and Post processing**: Reproducible Routing Rituals (RRR)
  - Written in Python3, leverages various pip packages
  - https://github.com/c-h-david/rrr
- **Sustainable DevOps**
  - Continuous Integration with Travis CI
  - Continuous Deployment with Docker Hub
  - Runs on laptop, desktop, cloud, or HPC
  - Inputs and outputs are generally netCDF or CSV. Example inputs/outputs available https://doi.org/10.5281/zenodo.3688690

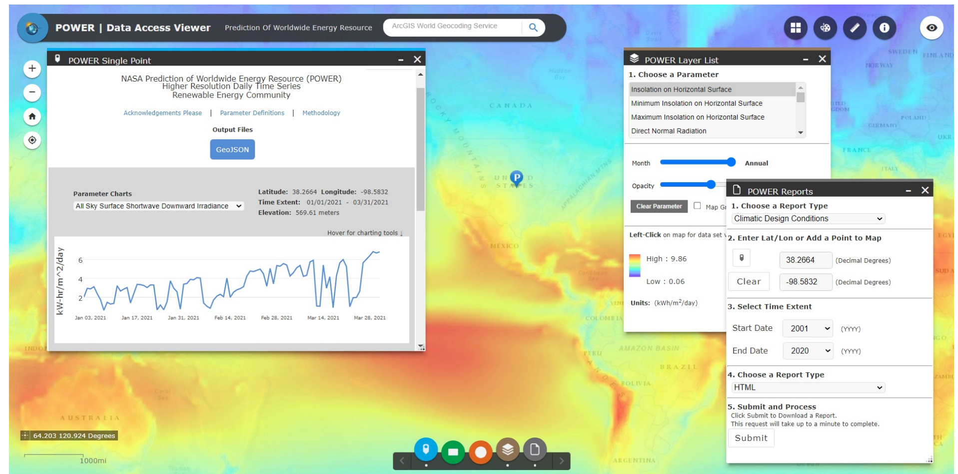


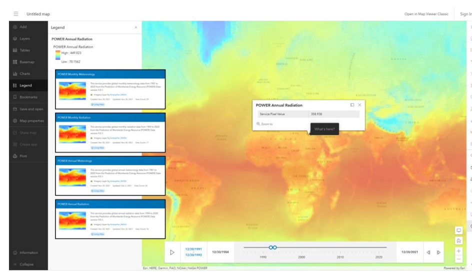
National Aeronautics and Space Administration

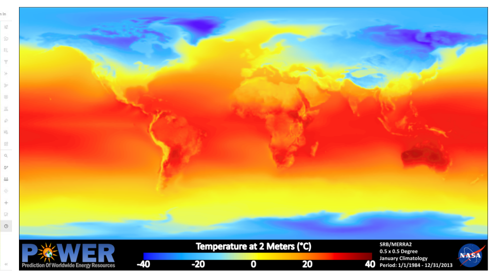
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

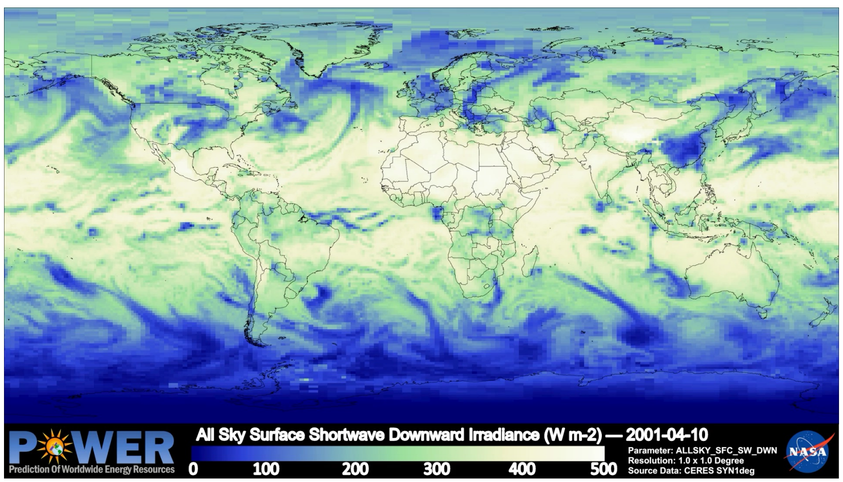
Initial Subsystem Validation: POWER

- Providing key input parameters to the hub in optimized data formats for use in the LIS model.
- Key parameters include
  - Surface meteorological properties
  - Downward surface radiative fluxes (both the solar and thermal infrared wavelengths). The data sources for these latter products is the CERES (Clouds and Earth's Radiant Energy System) mission.





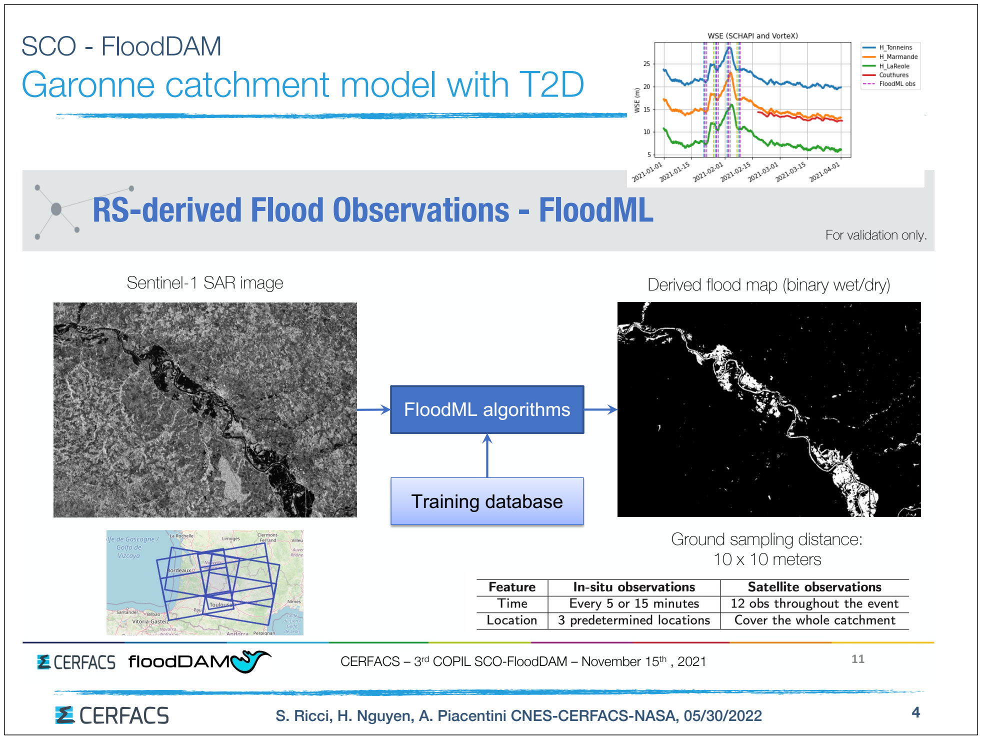


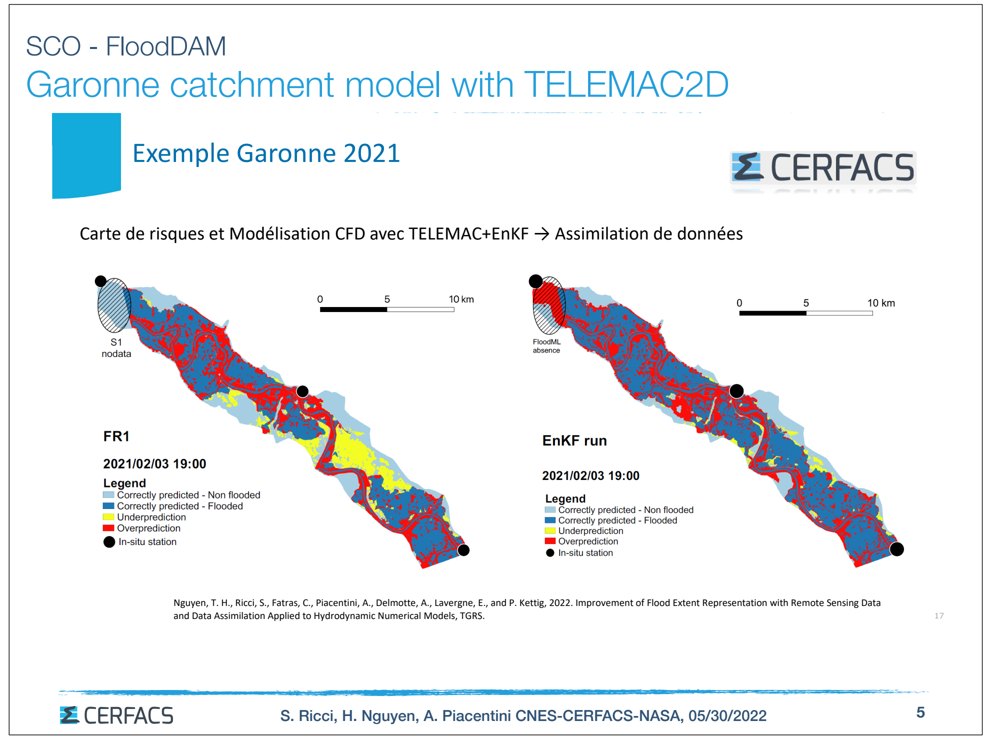


National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

SCO's FloodDAM's AI-based Risk Map



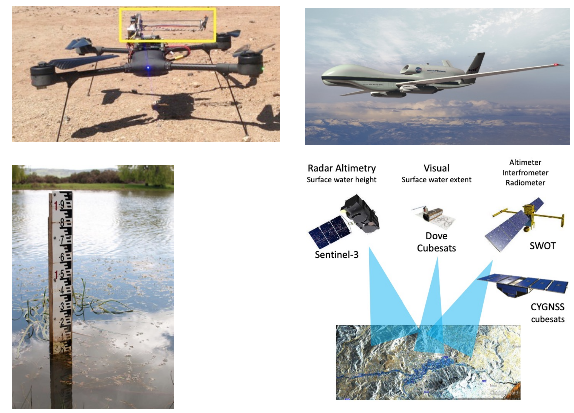


National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Actionable Predictions

- Model-driven scenario-based predictions
- Model-driven
  - Data acquisition
  - Data fusion
  - Value-added data processing
- Focuses on relevance
  - Data availability
  - Analysis and visualization capabilities



Acquire Observation and Analysis


Decision Support and Science Planning

National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

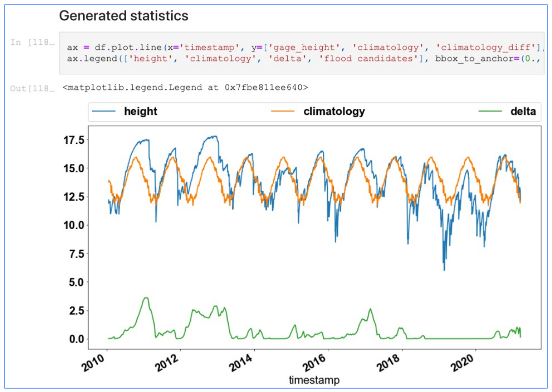
2017 Mississippi Flood, Steele Missouri

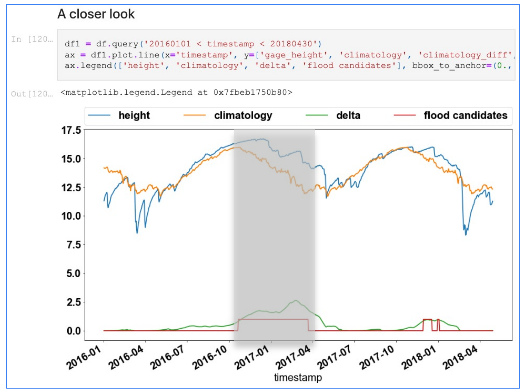
Integrate JPL-CNES Statistical Analysis, ML Prediction, and Advanced Image Processing

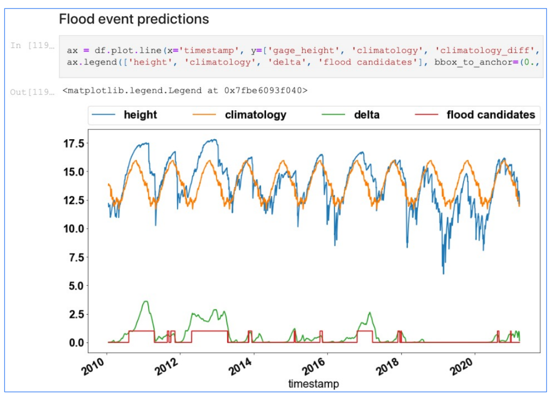
Huang, T., A. Altinok, N.T. Chung, J. Hausman, C.M. Oaida, S. Shah, Z.M. Taylor, S Baillarin, G. Blanchet, P. Kettig, and C. Taillan, "Distributed Machine Learning and Data Fusion for Flood Detection and Monitoring," 2020 AGU Fall Meeting, December 16, 2020.

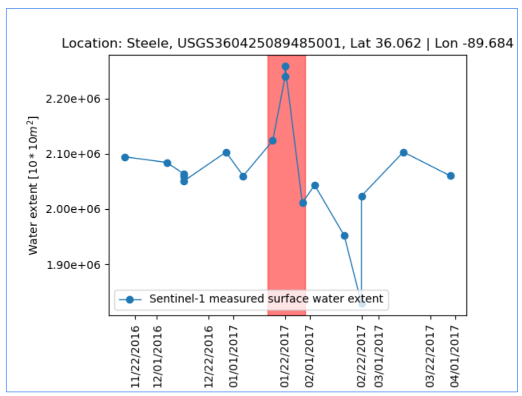
Model Prediction

USGS River Data



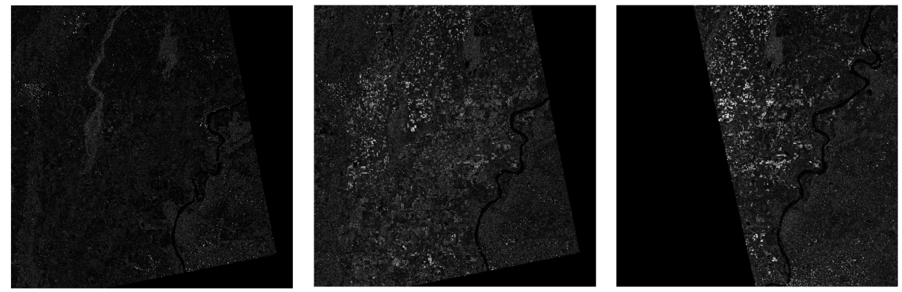


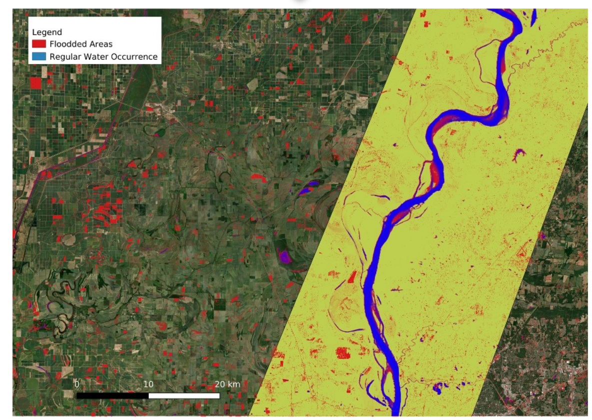




Corresponding Flood Map

Sentinel 1A





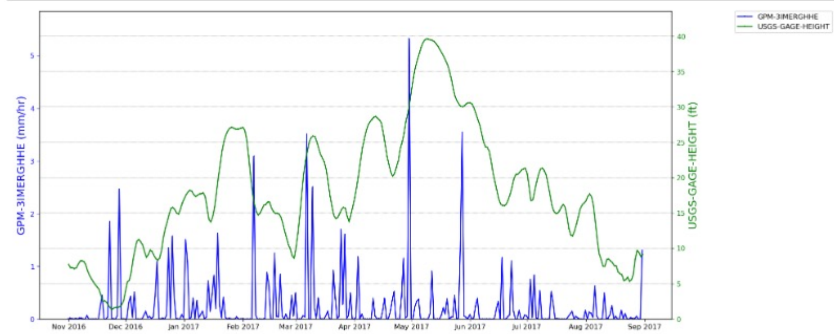
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

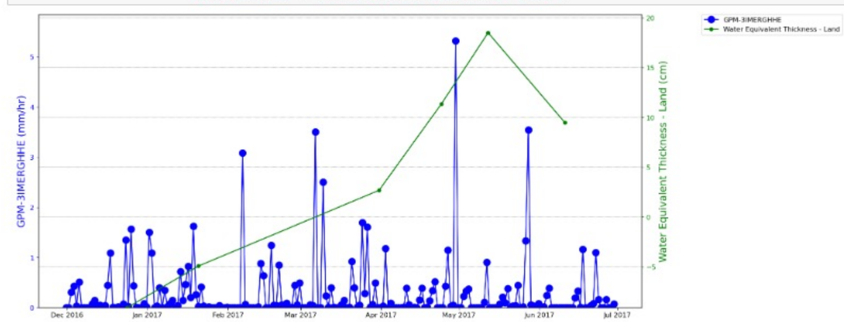
Initial Subsystem Validation: Apache SDAP and OnEarth

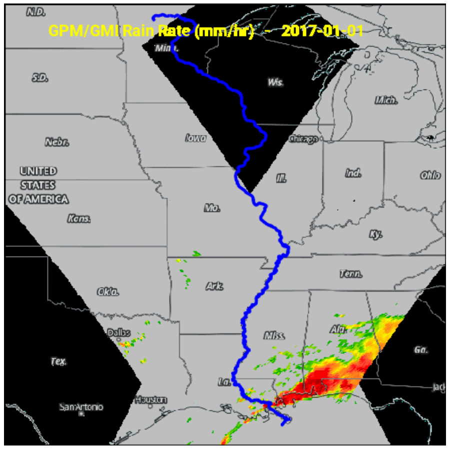
- Harmonization and Analysis
  - Onboarded sample data
    - GPM
    - Grace-FO (Federated)
    - LIS Output
    - USGS Gage Height
  - Validated using simple timeseries analysis
- Visualization
  - Generated tile-based visualization using GPM data
  - Developed tile-based animation on jupyter notebook

```
In [105]: show_plot_two_series(gpm_time, usgs_time, gpm_means, usgs_mean, "GPM-3IMERGHHE (mm/hr)", "USGS-GAGE-HEIGHT (ft)",
          "GPM-3IMERGHHE", "USGS-GAGE-HEIGHT")
```




```
In [141]: show_plot_two_series(gpm_time, grace_time, gpm_means, grace_means, "GPM-3IMERGHHE (mm/hr)",
          "Water Equivalent Thickness - Land (cm)",
          "GPM-3IMERGHHE", "Water Equivalent Thickness - Land")
```





GPM over Mississippi River

Initial Subsystem Validation: LIS output in SDAP and OnEarth

- Validation using sample LIS Soil Evaporation field using Jupyter
  - Tile visualization and animation
  - Time series analysis and data subsetting