Professional Open-Source Framework for Earth System Digital Twins and Applications

Clearance: CL#23-1745

© 2023. All rights reserved. California Institute of Technology. Government sponsorship acknowledged. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsements by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

Thomas Huang
NASA Jet Propulsion Laboratory
California Institute of Technology
Agenda ID: 2023.04.18_11.00
WGISS-55
Cordoba, Argentina (CONAE)
18-20 April 2023
Big Data is about being Smarter about Data
Agility | Relevancy | Sustainability

NASA Earth Science Archive Growth

AVAILABLE NOW!

Big Data Analytics in Earth, Atmospheric, and Ocean Sciences

NASA Earth Science Archive Growth
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
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

Decision Support and Science Planning

- 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
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 dynamic data acquisition, long-term prediction, data classification, process orchestration and management, etc.

Bridges the Physical Environment and its Virtual Representation
A community of analytic services that harmonizes data, tools and computational resources to permit the research community to focus on the investigation through common, portable web API.
IDEAS Teams

**Hydrology, Flood Prediction, and Analysis**
Partnership between NASA and the CNES-led Space for Climate Observatory (SCO)’s FloodDAM-DT effort

**NASA JPL:** Thomas Huang, Megan Bull (intern), Cedric David, Gary Doran, Jason Kang, Grace Llewellyn, Kevin Marlis, Stepheyne Perez, Wai (William) Phyo, Catalina M. Oaida, and Joe T. Roberts

**NASA GSFC:** Sujay V. Kumar and Nishan Biswas

**NASA LaRC:** Paul Stackhouse, David Borges, Madison P. Broddle, and Bradley MacPherson

**CNES:** Simon Baillarin, Lerre Benjamin, Frederic Bretar Gwendoline Blanchet, Peter Kettig, Raquel Rodriguez Suquet, and Lonjou Vincent

**CERFACS:** Sophie Ricci, Thanh-Huy Nguyen, and Andrea Piacentini

**Collecte Localization Satellites (CLS):** Christophe Fatras, Sylvain Brunato, and Eric Guzzonato

**QuanCube:** Alice Froidevaux, Antoine Guiot, Thanh-Long Huynh, and Romane Raynal

**Vortex.io:** Guillaume Valladeau and Jean-Christophe Poisson

**Wildland Fire, Air Quality, and Health Impact**
Partnership with NASA’s MAIA Mission, National Institute of Health, and City of Los Angeles

**NASA JPL:** Thomas Huang, Nga Chung, David Diner, Gary Doran, Sina Hasheminassab, Sarah Hallam (intern), Jason Kang, Olga Kalashnikova, Kyo Lee, Grace Llewellyn, Thomas Loubrieu, Kevin Marlis, Jessica Neu, Joe T. Roberts, and David Schimel

**City of Los Angeles:** Jeanne Holm, and Dawn Comer

**CSU Los Angeles:** Mohammad Pourhomayoun, and Pratyush Muthukumar

**Howard University:** Joseph Wilkins and Jonathan Barnes

**Washington University:** Randall Martin

**University of Colorado:** Daven Henze
Earth System Digital Twins for Hydrology, Flood Detection, and Analysis
IDEAS: Water Cycle Application

- **Resource Interface** - evaluate/extend emerging DT interface standards
- **Resource Discovery** - central resources registry
- **Resource Orchestration** - provide application-specific resource orchestration
- **Analytics Center** - Apache SDAP for data aggregation, harmonization, and analysis-optimized data services
- **New Observing Strategies** - event-based and ML-driven data acquisition and integration with FloodDAM-DT
- **Multiphysics, Multiscale, and Probabilistic Models** - LIS, RAPID, POWER, and FloodDAM-DT
- **Machine Learning and Deep Learning** - models for flood detection and prediction, data classification, services coordination, etc.
Federated Digital Twins between NASA and CNES

- Establish federated digital twins solution between the NASA ESTO/AIST's Integrated Digital Earth Analysis System (IDEAS) (Huang/JPL) and the Space for Climate Observatory (SCO) FloodDAM-DT (Rodriguez-Suquel/CNES)
- NASA AIST IDEAS is an open-source Earth System Digital Twins (ESDT) framework
- The collaboration focuses on establishing DT-powered flood alert systems, analysis, and risk maps on local and global scales
Built for Scenario-based Analysis

- Using the latest observation and analysis to drive model predictions
- Decision support
- On-demand data and analysis acquisition
- Future instrument scheduling and tasking

Observations & Analysis
Apache Science Data Analytics Platform (SDAP)
Land Surface Model
Land Information System (LIS) - NoahMP | Variable Infiltration Capability (VIC) | Catchment Land Surface Model (CLSM) | Noah
River Routing and Discharge Model
Routing Application for Parallel computation of Discharge (RAPID)
Flood Risk Impact
Water Level Mapping

Mississippi River USGS Streamflow and RAPID Discharge Land Surface Model Comparison

Mississippi Flood Event in 2018 – 2019
Comparing river discharge with USGS stream gauges

2021-02 Flood Event in Garonne River
Comparing river discharge with micro-station data
Bringing Observations and Models Together
2021-03 through 2021-12 in Garonne

NoahMP Average Surface Runoff

VIC Runoff and GPM Precipitation (normalized)

Vortex.io River Height

NoahMP Average Subsurface Runoff

RAPID Discharge from different Land Surface Models

Telemac3D Water Elevation
Minimize storage and computation need for pre-staging different in-situ sensor data. Use real-time ML to predict which stream gauges will be most useful for analysis

- **Precipitation-Only Approach**: use GPM data and ML model to predict daily peaks in discharge
- Random Forest model trained on 2,195 gages over 2 years, totaling 2.2 M examples, from midwestern US

- **Incorporating Stream Network**: use MERIT basin/reach database to model propagation of flow during flooding events
- Long Short-Term Memory (LSTM) neural network trained to forecast 6-24 hours into the future for each sensor given upstream readings
Demonstrates the latest IDEAS API and capabilities
- STAC – Data search and metadata
- Data access – satellite, in-situ, and models
- Interactive, harmonized data analytic capabilities
- Visualizations – Tile WMS and on-demand animation generation

IDEAS-Powered Flood Notebook
DT Powered VR – Immersive Science
SCO FloodDAM-DT Subsystems

- **VorteX.io**: Micro-stations
- **CERFACS**: Water Level Map
- **CNES and CLS**: Rapid Flood Mapping
- **QuantCube**: Financial Risk Map
- **CNES and CS**: Hydroweb platform and FloodDAM-DT integration
SCO FloodDAM-DT Pipeline

1- Water level measurement

2- Water height anomaly detection

3- Flood rapid mapping from EO imagery

4- Modelling

5- Visualization Platform

6- Risk map and Use Cases

- Financial institution
- Insurers
- Real estate funds
- Etc...

Standardize resource interfaces using emerging DT specifications (OGC standards)
• Cooperation IDEAS & SCO FloodDAM-DT

- CNES-AIST Kickoff Meeting: 18/2/2022
- Kick-Off JPL and French partners: 30/5/2022
- Interface & service specifications, selection of flood events: 2/12/2022
- Pre-demonstration between IDEAS services and FloodDAM: 1/6/2023
- End-to-End demonstration on selected catchments: 1/12/2023
- FloodDAM complete integration in Hydrowebnext Platform: 1/6/2024

- In-situ stations location & data access: 1/9/2022 - 1/6/2024
- FloodDAM-DT: 3/1/2022 - 1/6/2024
Earth System Digital Twins for Wildland Fire, Air Quality, and Health Impact
Big Picture Up Front

Improve usability of science data for air quality analysis and prediction

Leverage advanced assimilation of numerical and AI models to Improve Decision Making

Develop sustainable technology solution for sustainable science

“Characterize, understand, and improve the quality of air in urban areas across the planet.” – Jeanne Holm, Deputy Mayor for Budget and Innovation at LA
Data Harmonization
2017 – 2018 California Wildfire Season – Interactive Analyze and Visualize (PM$_{2.5}$, O$_3$, NO, CH$_4$, TLML, PBL, etc.)
Air Quality Data Analysis Tool
Powered by the AQ ACF Platform

- Interactive analysis and visualization of
  - Satellite Observations
  - Model for atmospheric composition
  - IOT and in-situ sensors
Analyze In-Situ Data
Example: PM$_{2.5}$ and Black Carbon from July 4$^{th}$ Fireworks

Dynamic retrieval of in-situ measurements
PM10, PM2.5, BC, CO, NO2, O3, AOD, etc.

Source: San Gabriel Valley Tribute

Compare 2021 and 2022 emissions

NASA Jet Propulsion Laboratory (JPL), Pasadena, CA

Source: San Gabriel Valley Tribune
Demonstrates the latest IDEAS API for air quality analysis

- STAC – Data search and metadata
- Data access – satellite, in-situ, and models
- Interactive, harmonized data analytic capabilities
- Visualizations – Tile WMS and on-demand animation generation

Air Quality Notebook
https://github.com/EarthDigitalTwin/IDEAS-notebooks/blob/main/AirQuality_Demo.ipynb
TROPOMI Global Daily Product (O$_3$, CH$_4$, NO$_2$, SO$_2$, and CO) Generation
Powers Value-added Product Generation

Read individual L2 swath files

Remap the data into HEALPix pixels at various spatial resolutions (6 km, 12 km, 24 km, …)

Global grid and generate a daily mean L3 dataset by averaging valid retrievals

Global NO$_2$ Total Column
2020-09-01 – 2020-09-03
California EV Mandate 2035 (a.k.a. CAL2035)

- The state of California plans to ban the sale of new gasoline-powered cars by 2035.
- Requires 35% of new vehicles sold in CA to be electric by 2026.
- Will increase to 68% in 2030 and 100% in 2035.
- These actions are estimated to achieve a more than 35% reduction in greenhouse gas emissions and an 80% improvement in NOx emissions from cars.
- Questions
  - How does air pollution in CA respond to changing vehicle emissions, and what are the impacts on human health and exposure?
  - How can air quality modeling data and remote sensing observations be used to present answers to these types of questions?
ML-Driven Data Acquisition

Hydrology, Flood Detection, and Analysis

Inputs:
Precipitation Data, [soil moisture, snow melt, …]

Model:
Predict propagation through river network using MERIT basins/reach database using LSTMs

Drives Decision:
Which in-situ stream gauge readings to ingest?

Wildfire, Air Quality, and Health Impact

Inputs:
MODIS, VIIRS, Wildfire Risk Index

Model:
Predict wildfire growth and pollutant propagation through atmosphere using (e.g.) GEOS-CF

Drives Decision:
Which in-situ air quality readings to ingest?
Future: MAIA Instrument Retargeting
Multi-Angle Imager for Aerosols (MAIA) and Surface Biology and Geology (SBG)

Exposure timescale

Health outcomes

Acute (days to weeks)
Emergency room visits, premature deaths

Subchronic (months)
Adverse birth outcomes, pregnancy complications

Chronic (years)
Cardiovascular and respiratory diseases

NOS – To simulate retasking of NASA’s Multi-Angle Imager for Aerosols (MAIA)

SBG’s fire detection and fire radiative power data for fire risk, detection, and analysis
Open-Source Science and Community Collaboration

- Partnership with Apache Software Foundation
- Embrace open-source software
- Evolve the technology through community contributions
- 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
Conclusion - Big Data is about being Smarter with Data

- Building toward Earth System Digital Twin opportunity to define and develop reusable, open-source Earth System Digital Twins framework
- Leverage, improve, and define community standards to promote interoperability
- It is about delivering professional quality open-source platform that addresses - Agility | Relevancy | Sustainability
- Enables end-to-end data and computation architecture, and the total cost of ownership
- Start with system architecture aiming for simple interfaces and information model - from generalization to specialization
- MUST team with Science Champions
- Successful big data platform needs to be ready for multi-computing, multi-data-center, and multi-agency

Thomas Huang
thomas.huang@jpl.nasa.gov
NASA Jet Propulsion Laboratory
California Institute of Technology

DARE MIGHTY THINGS
TOGETHER!
BACKUP
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)
• **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
• 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.
SCO FloodDAM-DT – catchment sites

France Catchment: **Garonne river**

- La Réole
- Sainte-Bazéille
- Couthures
- Marmande
- Tonneins

Legend:
- Vigicrue (national network) in-situ stations
- VorteX.io micro-stations
- EDF’s Telemac-2D model domain

USA Catchment: **Ohio River**

- Before (2018/02/11)
- Flooded (2018/02/23)

Legend:
- USGS streamgage stations
- Selected ROI
- Analyze live micro-stations data collected by VorteX.io
SCO FloodDAM-DT – Flood Mapping

- Rapid Flood mapping generation from SAR and optical data over the world
- Developed by CLS (Collecte Localisation Satellite) Group in France

**Washington State and Canada flood event 2021-11-16**

**Garonne Marmandaise flood event 2021-02-03**
Hydrodynamic model: Telemac 2D
FR: Free run simulation
IGDA: FR simulation with Data assimilation of in-situ and remote sensing observations

Publications:

Clear improvement of simulated flood extent maps using Data Assimilation of in-situ and RS data.
SCO FloodDAM-DT – Financial Risk Map

- Environmental Intelligence Platform and API
- Real time product on financial risk estimation and risk map generation of flooding for different type of assets

Flood risk map from FloodDAM-DT data combination with:
- Geolocation data of physical assets
- Geolocated Social media data
• FloodDAM-DT processing chain will be integrated in the hydrological platform Hydroweb.next
• CS group responsible of the FloodDAM integration, inter-operability and interfaces

Time series of water levels in the rivers and lakes around the world processed from satellite altimetry
Improvement of Flood Rapid mapping algorithm from SAR and optical data
Working on interfaces and interoperability
Chaining RAPID hydrological model with Telemac 2D hydrodynamic model
Working on the global financial physical risk model
Preparing future micro-station location and installation aux USA