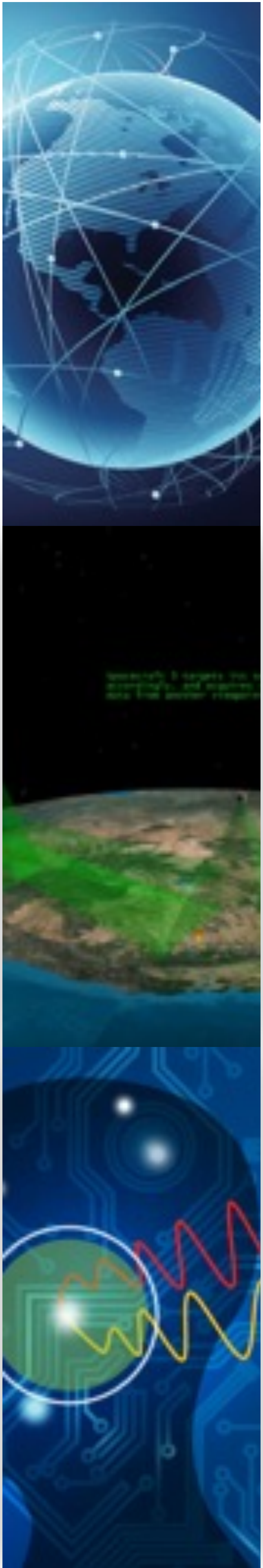


Advanced Information Systems Technology (AIST)

2023 Annual Reviews

Earth System Digital Twins (ESDT)

June 23, 2023

ESTO
Earth Science Technology Office

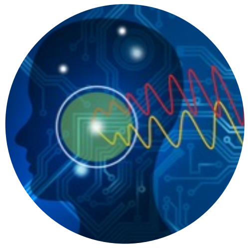
AIST Grouped Annual Reviews
ESDT Focus Area
Friday June 23, 2023

The Beckman Institute, Room 115
California Institute of Technology
1200 E. California Blvd.
Pasadena, California 91125

Time	End	Duration	Project #	PI (presenting)	Short Title
8:00 AM	8:20 AM	0:20	**Intro/Welcome**		
8:20 AM	8:45 AM	0:25	AIST-21-0003	Pelissier	TerraHydro
8:45 AM	9:10 AM	0:25	AIST-21-0031	Allen	Pixels for public health
9:10 AM	9:35 AM	0:25	AIST-21-0107	Da Silva	ACF for Earth System Observatory
9:35 AM	10:00 AM	0:25	AIST-21-0056	Bindlish	DT Infrastructure for Agriculture
10:00 AM	10:20 AM	0:20	**BREAK**		
10:20 AM	10:45 AM	0:25	AIST-21-0024	Keller	Ensemble prediction for atmospheric composition
10:45 AM	11:10 AM	0:25	AIST-18-0011	Martin	High-performance GEOS-Chem model
11:10 AM	11:35 AM	0:25	AIST-21-0063	Clune (Da Silva)	Global cloud-resolving OSSE framework
11:35 AM	12:00 PM	0:25	AIST-21-0052	Grubb	VALIXR: Extended Reality
12:00 PM	1:00 PM	1:00	**LUNCH**		
1:00 PM	1:25 PM	0:25	AIST-21-0095	Malik	Reproducible containers for collaborative analytics
1:25 PM	1:50 PM	0:25	AIST-21-0082	Katzfuss	Scalable UQ for Earth-system models
1:50 PM	2:15 PM	0:25	AIST-21-0012	Susiluoto	Kernel Flows
2:15 PM	2:40 PM	0:25	AIST-21-0091	Gray	Prototype DT of Air-Sea interactions
2:40 PM	3:05 PM	0:25	AIST-18-0099	Holm	Predicting what we Breathe
3:05 PM	3:25 PM	0:20	**BREAK**		
3:25 PM	3:50 PM	0:25	AIST-QRS-22-0001	Huang	IDEAS: open source ESDT framework
3:50 PM	4:15 PM	0:25	AIST-21-0018	Huang	Fire Alarm
4:15 PM	4:40 PM	0:25	AIST-21-0032	S. Lee	EcoPro
4:40 PM	5:00 PM	0:20	**Wrap-up/Adjourn**		

NASA ESTO
Advanced Information Systems Technology (AIST)

ESTO AIST

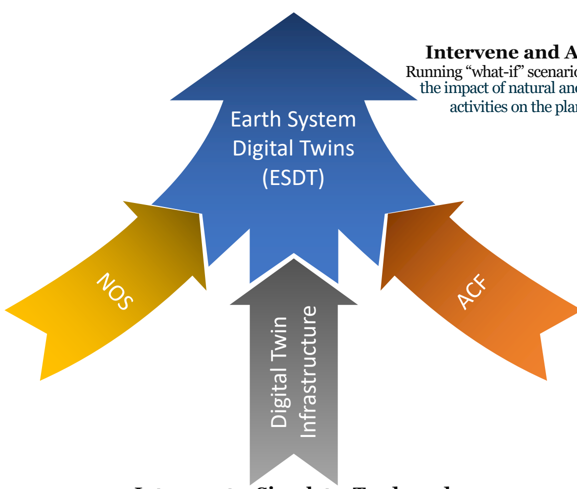

Earth Systems Digital Twins (ESDT)
Annual Grouped Technical Reviews

Jacqueline Le Moigne
June 23, 2023

Three AIST Thrusts

ESTO AIST

Intervene and Assess
Running "what-if" scenarios to assess the impact of natural and human activities on the planet.



Observe, Target and Coordinate
Edge and on-the-ground intelligent planning, evaluating, coordinating and operating collections of diverse and distributed observing assets

Fuse, Analyze, Share and Collaborate
Simplify access to diverse and large amounts of data, analytics & modeling tools and advanced computational resources for collaborative science

Interrogate, Simulate, Trade and Visualize
Robust tools for interrogating, assessing uncertainties & causality, and for visualization, leveraging diverse data, models and products

NOS = Novel Observing Strategies

ACF = Analytic Collaborative Frameworks

Earth System Digital Twins Components

ESTO
AIST

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.

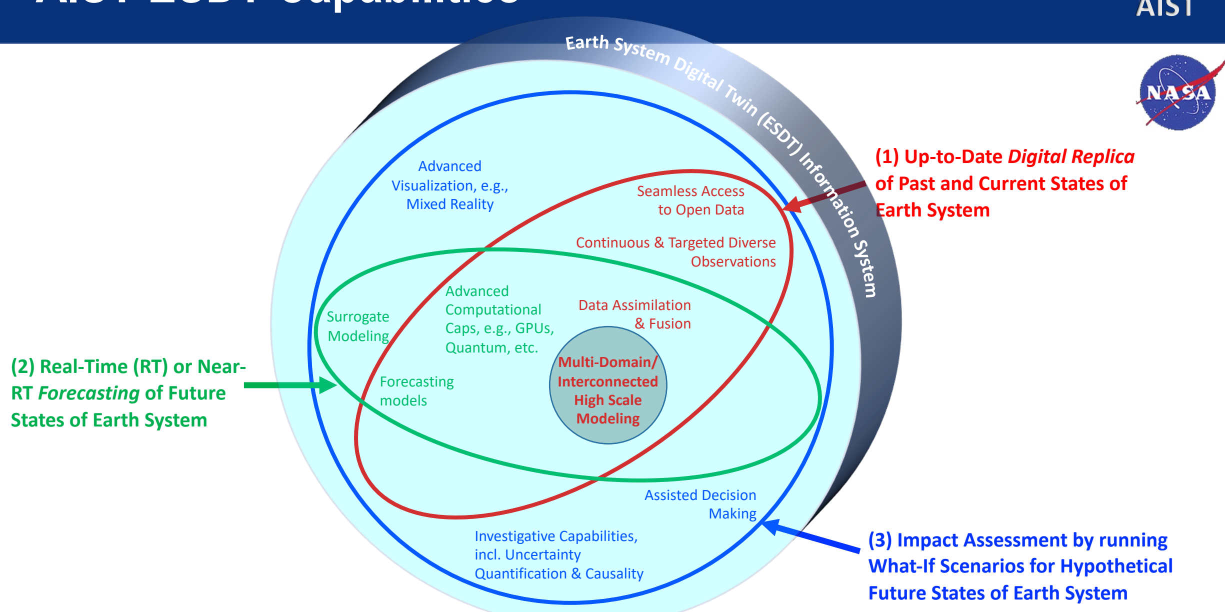
Impact Assessment . . . What if?

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

ESTO
AIST



ESTO AIST ESDT Activities Summary

- **AIST defines an Earth System Digital Twin (ESDT) as a dynamic and interactive information system that**:
  1. Provides a ***digital replica*** of the past and current states of the Earth or Earth system: *What-Now*
  2. Allows for computing ***forecasts*** of future states under nominal assumptions and based on the current replica: *What-Next*
  3. Offers the capability to ***investigate many hypothetical scenarios*** under varying impact assumptions: *What-If*.

- **AIST-21 Solicitation, first US government Solicitation requesting Digital Twins Technology for Earth Science:**
  - **New ESDT thrust** building on and advancing previously AIST-funded technology in ACF, NOS and Machine Learning (ML)
  - **14 AIST-funded ESDT projects (2020 – 2023)** focusing on developing:
    - Underlying analytic capabilities to build Digital Replicas
    - Novel ESDT infrastructure technologies
    - Surrogate modeling and ML emulators
    - Preliminary prototypes including interconnected modeling

- **AIST ESDT Workshop (October 2022)**
  - Report online on AIST Website by end of June 2023
  - Defined 6 science use cases during/after Workshop

- **International Coordination towards Interoperability and Standards:**
  - AIST-CNES Collaborative Development of Flood ESDT Prototype (IDEAS/FloodDAM DT)
  - Joint Conference Sessions and Coordination Meetings with ESA & DestinE

Grouped Reviews Objectives

- Respond to **Annual ESTO AIST Reporting Requirements**
  - Technical Annual Reviews Grouped by Focus Areas
  - Individual Programmatic Reporting
- Establish **Relationship between Awardees**
  - Assess complementarity of various approaches and technologies in same AIST thrust
  - Investigate potential collaboration/coordination opportunities (potentially share algorithms, codes or ideas)
  - Investigate 3rd Optional Year teaming arrangements:
    - If proposed, optional 3rd Years – will be *selected* 18 Months after project start
    - For one of three purposes:
      1. Transition AIST technology to another Program or project
      2. Develop NOS-Testbed Concept and/or Demonstration
      3. ESDT Prototype
    - Not all proposed 3rd Years might be funded
    - Can be different than original proposal but no budget increase
    - Collaborative AIST Projects will be prioritized/encouraged (i.e., several AIST projects in a system-of-systems approach
- Introduce AIST Projects and PIs to **Broader Community**
  - Present AIST projects to NASA ESD Program Managers/Scientists and partner organizations
  - Facilitate technology infusions and knowledge transfer of AIST projects upon completion.
- **Review Needs** in terms of:
  - SMCE (NASA Science Managed Cloud Environment): AWS system access
  - ESIP: Project analysis to improve infusion and transition opportunities

ESIP Evaluation

AIST

Between 12 and 18 months in your project, you can request an **Assessment of Maturity by ESIP** ("Earth Science Information Partners")

- **No cost to the PIs**

- **Process:**
  1. ***Objectives Set up and Facilitation:***
     - ESIP provides access to the Earth Science community & feedback on your technology/product/tool
     - ESIP will work with PIs to set specific objectives, taking into consideration TRL
     - ESIP will facilitate evaluator calls, development of evaluation plan, communication with PIs
  2. ***Technical Exchange Meeting:***
     - PI team meets evaluators.
     - Big picture to backend… evaluators should have a solid understanding of the purpose and goals of technology
  3. ***Evaluation Period:***
     - ESIP coordinates evaluation process.
     - Evaluators meet regularly, requesting information from PIs when necessary.
  4. ***Final Report:***
     - ESIP works with evaluators to create final report to be shared with PIs & AIST.
     - Reports can be public upon PI request.

AIST

ESTO
Earth Science Technology Office

Building AI-based Earth Systems Models and the TERRAHydro Terrestrial Digital Twin

*Craig Pelissier (PI)
Grey Nearing (Co-PI)
Carlos Cruz
Kia Saeedi
Brandon Smith
Vanessa Valenti

*presenter

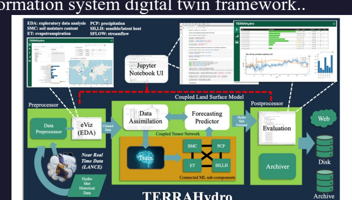
Terrestrial Environmental Rapid-Replication and Assimilation Hydrometeorological System (TERRAHydro)

PI: Craig Pelissier, Science Systems Applications Inc.

Objective

To develop a coupled water, energy, and vegetation ESDT (TERRAHydro) using tensor-based modeling that leverages the speed and accuracy of ML inference to provide unprecedented capabilities including (but not limited to) assimilation, rapid-reconfigurability, comprehensive scenario and What-if Analyses, and uncertainty quantification.

Technology Framework TERRAHydro deploys a modern, open-source, open-science Python-based information system digital twin framework..



Approach:

1. To couple the current state-of-the-art hydrometeorological ML Tensor Network models using 3 coupling strategies (direct coupling, shared model structure, and PDE learning).
2. To develop a modern Python-based information systems encapsulating the proposed land surface model that is open-source, cloud-ready, portable, and enables open-science.
3. To assess and demonstrate capabilities on the 2006-2010 Syrian drought and water storage changes in the Himalayan mountains.

CoIs: **Grey Nearing, Carlos Cruz, Brandon Smith, and Vanessa Valenti.**

Key Milestones

- **Software Milestone (year 1):** Basic functionality to enable training, validation, and coupling.
- **Software Milestone (year 2):** All Functionality completed for final year assessment + demonstration.
- **Software Milestone (year 3):** Cloud-ready complete ML workflow capabilities.
- **Technology Milestone (year 1):** All training data acquired and LSM sub-components tested.
- **Technology Milestone (year 2):** Assimilation, coupling, physics-informed methodologies done.
- **Technology Milestone (year 3):** Demonstrate and assess the power of the TERRAHyrdo.

$TRL_{in} = 3$

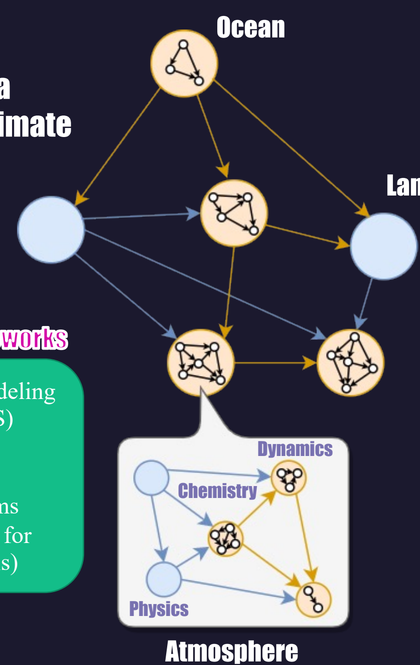
June, 2023

ESTO Earth Science Technology Office

The Structure of Earth Systems Models

- Composed of a coupled set of sub-systems or models of physical processes.
- Organizable into a hierarchical graph with components developed and maintained by domain experts across multiple organizations (e.g., government, academia, industry).
- Frameworks for coupling, organization, and operation of large ESMs/climate models are essential to developing and maintaining large open-source open-science systems.
- Existing frameworks mostly written in/for Fortran and parallelized with MPI for scalability on large CPU-based clusters.

A Diagram of a Hierarchical Climate Model



Some existing frameworks

- Earth Systems Modeling Framework (GEOS)
- MAPL (GEOS)
- NASA's Land Information Systems (operational driver for land surface models)

06/23/23

Adoption of AI-based Technology

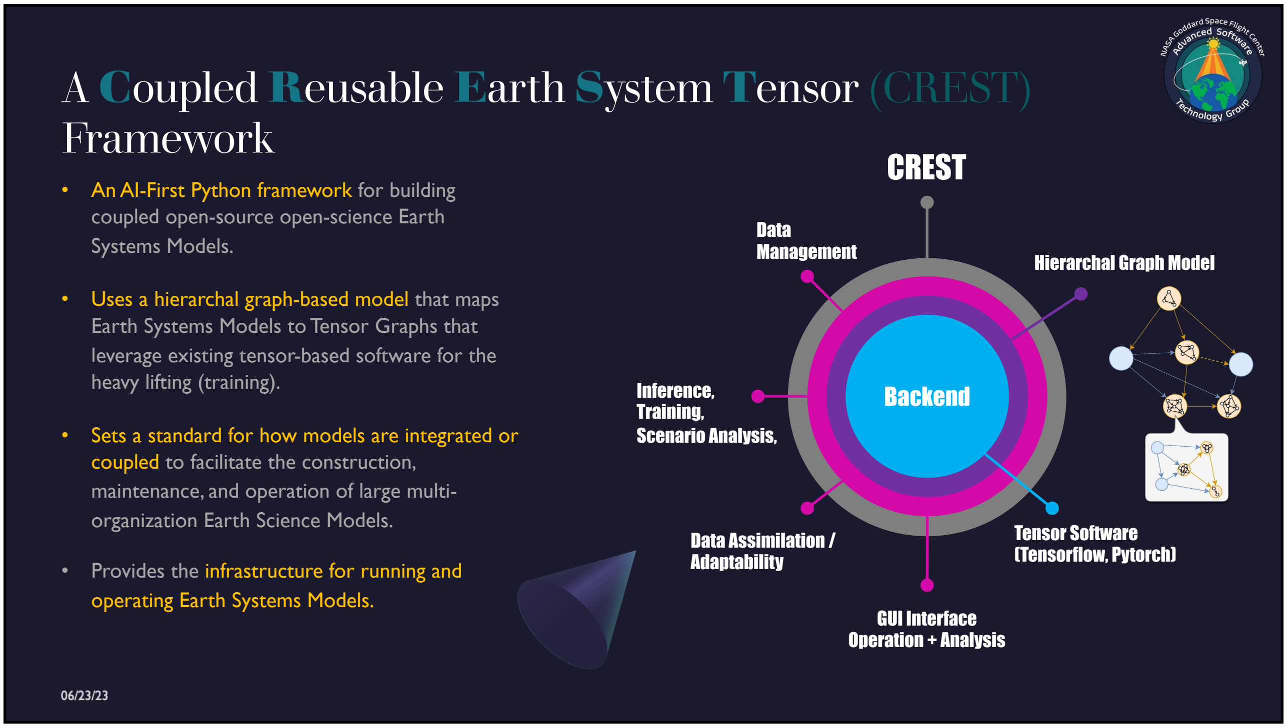
- Data-Driven models have emerged across many Earth Science domains and have been shown to be a powerful alternative to traditional modeling.
- Very few operational AI-based Earth Systems models exists today (one example is Google's flood forecasting system).
- Integration into existing modeling frameworks is a significant challenge to the adoption AI-based technologies.

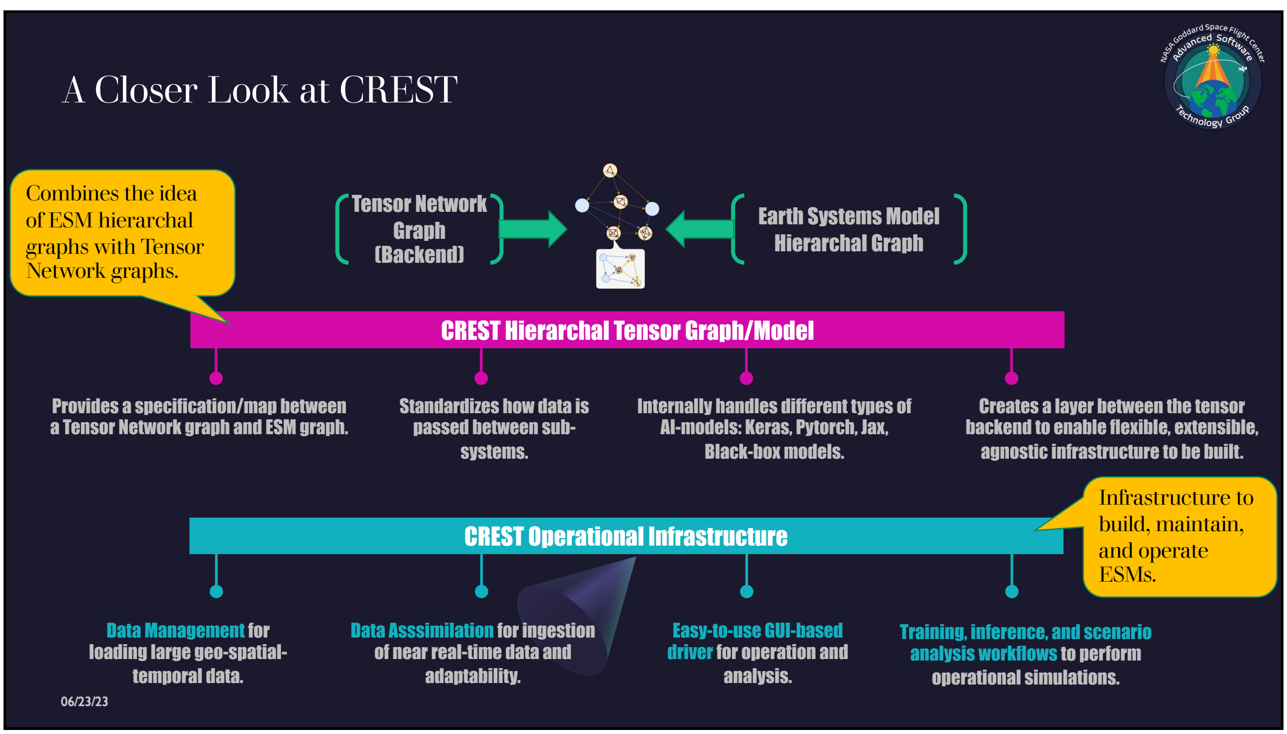
This is relevant today. AI technologies already exist that offer improved capabilities.

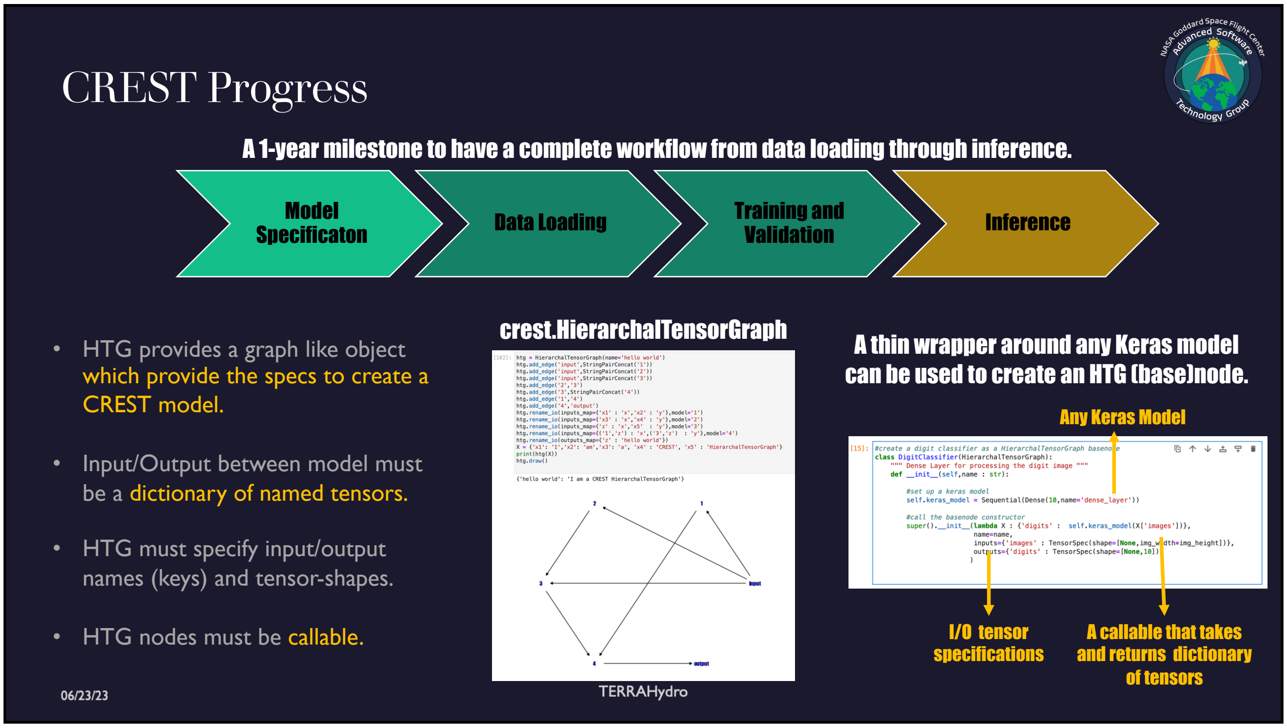
Some Barriers to Adoption

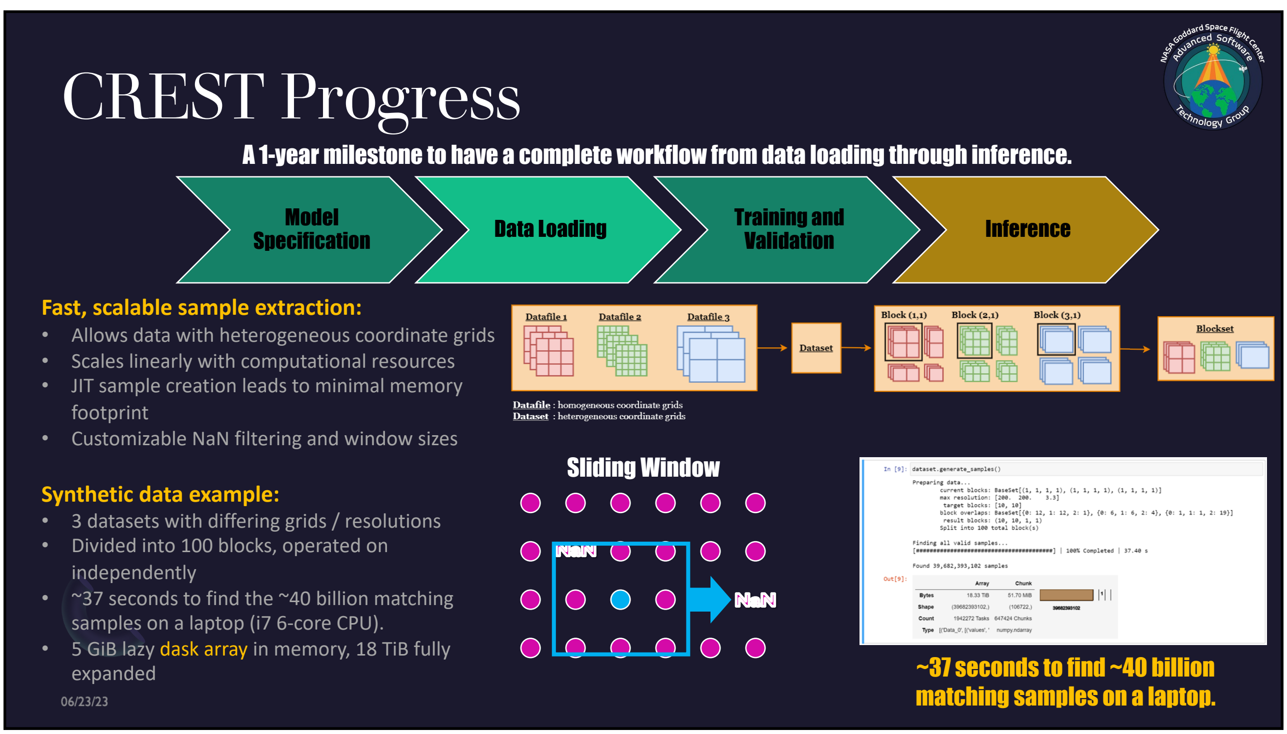
- Existing frameworks do not provide gradients to train fully integrated AI technologies.
- AI models are written (Python) using powerful tensor-based software which is not easily integrated with Fortran/MPI.
- AI-models use differently structured data (sequence-to-one sequence-to-sequence, image stacks/series)
- Assimilation techniques in AI can use tensor gradients techniques.

06/23/23


A Coupled Reusable Earth System Tensor (CREST) Framework
An AI-First Python framework for building coupled open-source open-science Earth Systems Models.
Uses a hierarchal graph-based model that maps Earth Systems Models to Tensor Graphs that leverage existing tensor-based software for the heavy lifting (training).
Sets a standard for how models are integrated or coupled to facilitate the construction, maintenance, and operation of large multi-organization Earth Science Models.
Provides the infrastructure for running and operating Earth Systems Models.
CREST
Data Management
Hierarchal Graph Model
Inference, Training, Scenario Analysis,
Backend
Data Assimilation / Adaptability
Tensor Software (Tensorflow, Pytorch)
GUI Interface Operation + Analysis
06/23/23


A Closer Look at CREST
Combines the idea of ESM hierarchal graphs with Tensor Network graphs.
Tensor Network Graph (Backend)
Earth Systems Model Hierarchal Graph
CREST Hierarchal Tensor Graph/Model
Provides a specification/map between a Tensor Network graph and ESM graph.
Standardizes how data is passed between sub-systems.
Internally handles different types of AI-models: Keras, Pytorch, Jax, Black-box models.
Creates a layer between the tensor backend to enable flexible, extensible, agnostic infrastructure to be built.
Infrastructure to build, maintain, and operate ESMs.
CREST Operational Infrastructure
Data Management for loading large geo-spatial-temporal data.
Data Asssimilation for ingestion of near real-time data and adaptability.
Easy-to-use GUI-based driver for operation and analysis.
Training, inference, and scenario analysis workflows to perform operational simulations.
06/23/23


CREST Progress
A 1-year milestone to have a complete workflow from data loading through inference.
Model Specificaton
Data Loading
Training and Validation
Inference
HTG provides a graph like object which provide the specs to create a CREST model.
Input/Output between model must be a dictionary of named tensors.
HTG must specify input/output names (keys) and tensor-shapes.
HTG nodes must be callable.
crest.HierarchalTensorGraph
A thin wrapper around any Keras model can be used to create an HTG (base)node.
Any Keras Model
I/O tensor specifications
A callable that takes and returns dictionary of tensors
06/23/23
TERRAHydro

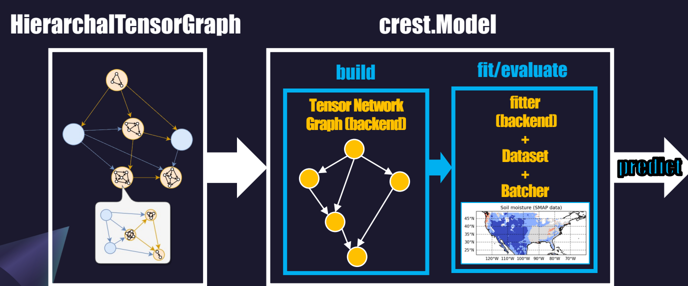

CREST Progress
A 1-year milestone to have a complete workflow from data loading through inference.
Model Specification
Data Loading
Training and Validation
Inference
Fast, scalable sample extraction:
Allows data with heterogeneous coordinate grids
Scales linearly with computational resources
JIT sample creation leads to minimal memory footprint
Customizable NaN filtering and window sizes
Synthetic data example:
3 datasets with differing grids / resolutions
Divided into 100 blocks, operated on independently
~37 seconds to find the ~40 billion matching samples on a laptop (i7 6-core CPU).
5 GiB lazy dask array in memory, 18 TiB fully expanded
Datafile 1
Datafile 2
Datafile 3
Dataset
Block (1,1)
Block (2,1)
Block (3,1)
Blockset
Datafile : homogeneous coordinate grids
Dataset : heterogeneous coordinate grids
Sliding Window
NaN
NaN
~37 seconds to find ~40 billion matching samples on a laptop.
06/23/23

CREST Progress

A 1-year milestone to have a complete workflow from data loading through inference.

Model Specification → Data Loading → Training and Validation → Inference

- The CREST Model translates an HTG into a tensor network graph using the backend and provides functionality for fitting, evaluation, and prediction.
- Currently, it can only build TensorFlow-based models.
- Provides:
  - build() translates the HTG into a tensor network graph.
  - fit() train the model
  - evaluate() returns the model loss.
  - predict() makes a prediction given input.



06/23/23

TERRAHydro

Traditional Modeling In CREST

- It is easier to express physical processes in Tensor Networks than integrate tensor networks in physical-based model.
- Any physical-based or process-based parameterized model can be written using tensor-based software.
- Some work on incorporating physical constraints within AI models.
- Some approaches to estimate gradients of black-box models which would allow wrapping of Fortran-based models into python for inclusion in CREST.
- Some efforts towards Fortran auto-diff.

Integrating traditional models in CREST is targeted for future development.

NEED GRADIENTS

Potential Approaches

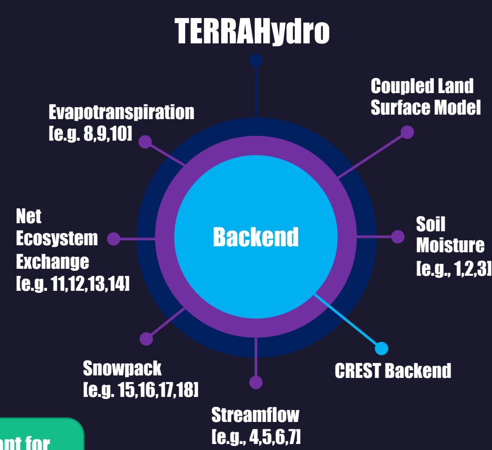
- Rewrite (Fortran) models in a tensor language.
- Leverage black-box gradient estimates.
- Development of a reliable auto-differentiator for Fortran.
- Apply physical constraints on AI models.

Emerging packages like JAX offer the use of numpy-like numerical packages that make this easier.

06/23/23

TERRAHydro

- Takes the next step in data-driven hydrology by coupling together 5 existing AI-based land surface components to assemble a land surface digital twin.
- Demonstrates an application of the CREST framework for assembling coupled AI-based Earth Systems Models --- guides the development.
- Enables new tensor-gradient-based data assimilation techniques in addition to traditional Bayesian approaches for near-real time ingestion of data.
- Computational efficiency and use of hardware accelerators enables rapid adaptability and scenario analysis beyond current capabilities.
- Potential for improved accuracy. References can be found on the last slide.

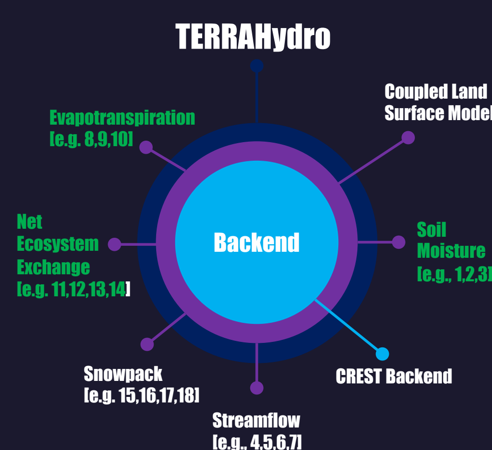


Important for Digital Twin capabilities

June, 2023 TERRAHydro

TERRAHydro Progress

- Soil Moisture, Evapotranspiration, Net Ecosystem Exchange (3 of 5) land surface components have been implemented and moderately validated.
- ESMWF (ERA5), Soil Moisture Active Passive (SMAP), and FluxNet data has been ingested for training.
- Large scale validation to reproduce paper results underway.
- A coupled Soil Moisture + Evapotranspiration model is underway with a target demonstration of coupled Soil Moisture + Evapotranspiration + Net Ecosystem Exchange model at the 18-month demo



June, 2023 12

Summary/Plan Forward

CREST

- Nearly completed the 1-year milestone of a complete pipeline from data loading through inference.
- Implemented the data loader, model specification (HTG), and model building and training.
- Can now build TERRAHydro with TensorFlow.
- Finish 1-year milestone and continue to add features and refinements as needed.
- Expand HTG to enable any node that provides a gradient and trainable parameters.
- NASA public license to share with external collaborators underway.
- Data assimilator and GUI-driver.

TERRAHYDRO

- 3 out of 5 model moderately validated (SM, ET, NEE).
- FluxNet, ERA5, and SMAP data ingested.
- Prototype ET+SM model underway.
- Fully validate SM, ET, and NEE (longer runs that produce relatively accurate maps).
- Build and moderately validate a SM + ET + NEE in preparation for 18-month demo.
- Fully validate (produce maps) using SM + ET + NEE.
- Data assimilation, validation of SF and SP, and moderately validated model with all 5 components in preparation for year-3 demos.

06/23/23

Timeline

- September, 2022: Project kick-off.
- 1st year milestone: End-to-end capabilities in CREST and a simplified coupled model with SM+NEE+ET.
- 18 month demo: TERRAHydro retrospective demonstration of a coupled land surface model for a hydrological event.
- 2nd year milestone: TERRAHydro coupled land surface model and assimilation completed along with all relevant CREST features for the third-year demo.
- 3rd year: Full scale retrospective demonstration including data assimilation and scenario analysis on relevant hydrological events.

06/23/23

Potential Infusion/Collaborations

Provide a summary of actual or potential infusions

- Infusion: The Google Flood Forecasting team is working to expand their operational, global AI-based flood forecasting model to include drought monitoring and S2S time horizons. Our objective is for CREST to support that expansion.
- Knowledge Transfer: The US National Water Model is being redesigned (a project called the Nextgen National Water Model), and the interoperable modeling framework used by that project (called the Basic Model Interface) does not support ML.

Provide a summary of actual or potential collaborations

- Co-PI Nearing is co-leading the development of an expanded land surface model at Google Research, and that team will leverage tools from the CREST project for that effort.
- Co-PI Nearing is working with the CIROH project (30 universities, 151 PIs including Nearing, $360M) to integrate ML into the Nextgen NWM. The Nextgen BMI has the same problems as ESMF in terms of supporting ML. Our role on that project is to advise about how to expand BMI to support ML.
- PI Craig Pelissier is working/attending meetings of a collaborative effort between CNES, NOAA and NASA to build a Coastal Zone Digital Twin and advising on the potential of leveraging CREST for that effort.

06/23/23

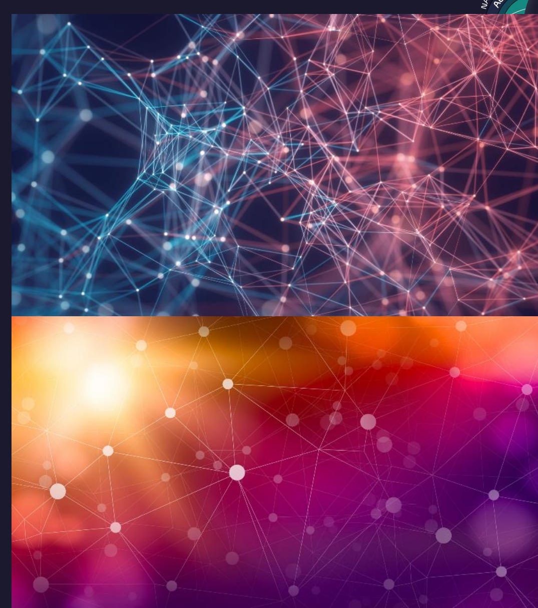
Thank You

Craig Pelissier, craig.s.pelissier@nasa.gov

Lead, Advanced Software Technology Group (ASTG)

NASA Goddard/SSAI

ASTG website



06/23/23

References

[1] Kuai Fang, Chaopeng Shen, and Daniel Kifer. Evaluating aleatoric and epistemic uncer- tainties of time series deep learning models for soil moisture predictions. arXiv preprint arXiv:1906.04595, 2019.
[2] Kuai Fang, Chaopeng Shen, Daniel Kifer, and Xiao Yang. Prolongation of smap to spa- tiotemporally seamless coverage of continental us using a deep learning neural network. Geophysical Research Letters, 44(21):11–030, 2017.
[3] O Sungmin and Rene Orth. Global soil moisture data derived through machine learning trained with in-situ measurements. Scientific Data, 8(1):1–14, 2021.
[4] Frederik Kratzert, Daniel Klotz, Mathew Herrnegger, Alden K Sampson, Sepp Hochre- iter, and Grey S Nearing. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. Water Resources Research, 55(12):11344–11354, 2019.
[5] Frederik Kratzert, Daniel Klotz, Claire Brenner, Karsten Schulz, and Mathew Herrneg- ger. Rainfall–runo↵ modelling using long short-term memory (lstm) networks. Hydrology and Earth System Sciences, 22(11):6005–6022, 2018.
[6] Frederik Kratzert, Daniel Klotz, Guy Shalev, Gu¨nter Klambauer, Sepp Hochreiter, and Grey Nearing. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences, 23(12):5089–5110, 2019.
[7] Dapeng Feng, Kuai Fang, and Chaopeng Shen. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. Water Resources Research, 56(9):e2019WR026793, 2020.
[8] Grey S Nearing, Benjamin L Ruddell, Martyn P Clark, Bart Nijssen, and Christa Peters- Lidard. Benchmarking and process diagnostics of land models. Journal of Hydromete- orology, 19(11):1835–1852, 2018.
[9] Wen Li Zhao, Pierre Gentine, Markus Reichstein, Yao Zhang, Sha Zhou, Yeqiang Wen, Changjie Lin, Xi Li, and Guo Yu Qiu. Physics-constrained machine learning of evapo- transpiration. Geophysical Research Letters, 46(24):14496–14507, 2019.
[10] Andrew Bennett and Bart Nijssen. Deep learned process parameterizations provide better representations of turbulent heat fluxes in hydrologic models. Water Resources Research, 57(5):e2020WR029328, 2021.
[11] Phuong Nguyen and Milton Halem. Machine learning for inferring co2 fluxes: The new metaphysics of neural nets. UMBC Faculty Collection, 2019.
[12] O Reitz, A Graf, M Schmidt, G Ketzler, and M Leuchner. Upscaling net ecosystem exchange over heterogeneous landscapes with machine learning. Journal of Geophysical Research: Biogeosciences, 126(2):e2020JG005814, 2021.
[13] Jianzhao Liu, Yunjiang Zuo, Nannan Wang, Fenghui Yuan, Xinhao Zhu, Lihua Zhang, Jingwei Zhang, Ying Sun, Ziyu Guo, Yuedong Guo, et al. Comparative analysis of two machine learning algorithms in predicting site-level net ecosystem exchange in major biomes. Remote Sensing, 13(12):2242, 2021.
[14] Ni Huang, Li Wang, Yuelin Zhang, Shuai Gao, and Zheng Niu. Estimating the net ecosystem exchange at global fluxnet sites using a random forest model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14:9826–9836, 2021.
[15] Ludovica De Gregorio, Mattia Callegari, Carlo Marin, Marc Zebisch, Lorenzo Bruzzone, Begu¨m Demir, Ulrich Strasser, Thomas Marke, Daniel Gu¨nther, Rudi Nadalet, et al. A novel data fusion technique for snow cover retrieval. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(8):2862–2877, 2019.
[16] Changyu Liu, Xiaodong Huang, Xubing Li, and Tiangang Liang. Modis fractional snow cover mapping using machine learning technology in a mountainous area. Remote Sensing, 12(6):962, 2020.
[17] Jinliang Hou, Chunlin Huang, Ying Zhang, Jifu Guo, and Juan Gu. Gap-filling of modis fractional snow cover products via non-local spatio-temporal filtering based on machine learning techniques. Remote Sensing, 11(1):90, 2019.
[18] Soni Yatheendradas and Sujay Kumar. A novel machine learning-based gap-filling of fine-resolution remotely sensed snow cover fraction data by combining downscaling and regression. Journal of Hydrometeorology, 2021.

Pixels for Public Health:

Analytic Collaborative Framework and Digital Twin to Enhance Coastal Resiliency of Vulnerable Populations in Hampton Roads, Virginia

Dr. Thomas Allen, PI, Old Dominion University

AIST-18/21-0031 Annual Technical Review

June 23, 2023

Team: Co-Is Navid Tahvildari, George McLeod, Heather Richter (Old Dominion University), Soenke Dangendorf (Tulane University), and Oz Yetkin (Analytical Mechanics Associates)

1

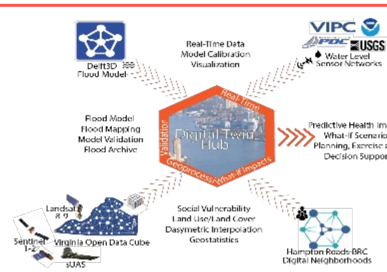
Pixels for Public Health: Digital Twin for Coastal Resilience of Vulnerable Populations

PI: Thomas Allen, Old Dominion University

Objective

Integrate multi-source, high-velocity data: Earth Observations, hydrodynamic models, and sensor networks to predict human health impacts of coastal flooding. Prototypes a coastal Digital Twin that integrates sea-level change, hydrodynamics, and extreme rainfall modeling and its impacts on an urban system.

- Optimize EO data streams, flood modeling, and IoT sensors for validation and model improvement.
- Align observations and predictions to decision support tools in short- and long-term scenarios for health and emergency management and sea level rise adaptation and planning.
- Develop EEJ impacts and metrics tor measuring social disparities of climate change impacts across the region.



Coastal Digital Twin Concept to Schematic

Approach

Utilize an analytical collaborative framework and Open Data Cube to replicate Hampton Roads coastal flooding.

Create a prototype Digital Twin to integrate flood models, sensor and satellite observations and analyze human health vulnerabilities from coastal flooding.

Collate and make interoperable STACs for flood models and sensor observation networks. Use Hub technology to visualize DT scenarios.

Code linkages of disparate data structures and harmonize their use for predictive analytics (risk exposure, accessibility models, health morbidities, health services and mitigation)

Create a platform for *what-if* scenario analysis, testing, and tabletop functional exercises for end-users

Key Milestones

• VA Data cube and product prototypes	12/22
• SLR modeling scenarios and probabilities	02/23
• Flood model domain, assessment, GIS workflow	04/23
• Flood sensor + IoT app integration	08/23
• Health impacts digital twin prototype	12/23
• Future SLR + flood scenarios implemented	09/24
Optional Year 3:	
• Tabletop exercise and sustainability plan	03/25

Co-Is/Partners: Tulane University, VIPC, NWS, Hampton Roads BRC

TRL_{in} = 7 **$TRL_{current}$ = 7.5**

Presentation Contents

3

Background and Objectives

Background: This project helps supports NASA R&A and Applications science goals:

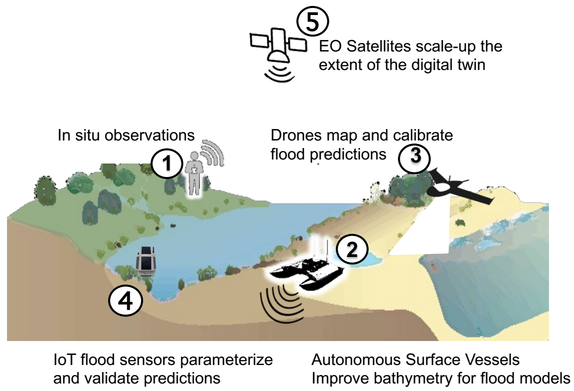
- **Climate Variability**: Improved understanding of sea-level acceleration and how to include it in future probabilistic scenarios with attribution of multiple forcings.
- **Weather**: Incorporate rainfall with tidal and storm surge flooding to characterize coastal compound flooding.
- **Disasters; Health & Air Quality; Water & Food, EEJ**:
  - NASA-Infused Digital Twin improve on the response, recovery and planning for acute to chronic coastal hazards
  - Improved characterization of population vulnerability and EEJ
  - Demonstrate scenario-based and what-if analysis DT applications

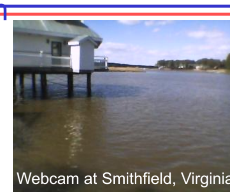
Objectives:

1) Develop and showcase the potential use of new Digital Twin technology to support coastal resilience.
2) Expand capabilities of the Virginia Open Data Cube to integrate multi-source IoT flood sensor data, flood model predictions, and Earth Observations.
3) Demonstrate a prototype DT with analytical capability of robustly replicating complex flood events in Hampton Roads, Virginia, while also enabling future what-if analyses applied to emergency management, public health, and equity and environmental justice.
4) Evaluate potential future advances for DT technology and its sustainable use.

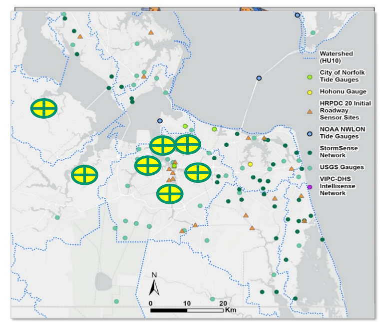
4

Inventorying Technology:
A variety of observing systems are being combined in the digital twin





Disparate flood sensors need to be combined.
Gaps in coverage need to be filled.



ESTO Earth Science Technology Office

Background / Objectives / Tech Advance

- ***Initial findings***
  - The VA Open Data Cube is developing new capabilities across IoT flood sensors, EO data, and flood prediction models.
  - We discerned more detail in the historical sea-level acceleration and are modeling with CMIP and future large ensemble models
  - Discovered unexpected point source pollution connectivity to tidal flooding in our hydro-connectivity analyses.
  - Generated new code for DataCube analysis (shared via Git and Hub)
  - Tested new data cube notebooks using OPERA DSWx, VIMS TideWatch flood model, six flood sensor networks, and other EO data cloud repositories.
  - Identified new opportunities for optimization of data streams (e.g. COG-format and geoprocessing, ingesting UAS data, and hosting cloud data services)

- ***Phasing Adjustments***
  - First-year phasing required slight budget adjustments; securing Tulane subaward with necessary resource to fund the upfront SLRise work.
  - Sensor integration and data cueb implementation were initially slow but recently apace.
  - Revised AMA and VMASC subcontracts to support additional AWS data cube work (tiling and data ingestation, future scalability, and front-end)

ESTO Earth Science Technology Office

Technology Development Results

- The Virginia Open DataCube (ODC) has been redesigned and stocked with ARD datasets, including new products and STAC
    - Assimilated multiple IoT flood sensor networks in a unified/common data structure for platforms via AWS: *StormSense* and Esri ArcGIS Hub dashboard
    - OPERA DSWx Landsat-Sentinel 2 Harmonized Water Extent
    - SWOT pending
    - Dashboard is in draft state for activation on the Hub
  - Geospatial Modeling has developed new algorithms to identify hydroconnected flood vulnerabilities (automated with LiDAR DEMs, USGS NHD, and regional stormwater and transportation infrastructure)
    - Hydro-correction models have performed successfully and are being regionally implemented, soon to be published to the Hub and datacube.
  - Hydrodynamic modeling is progressing with precipitation
  - VA OpenData Cube AWS registries are being compiled for integration
  - New product definitions for UAV and ASV autonomous data.

- Residual Project Risks and Mitigation
  - Departed Col A. Nielsen, onboarded AMA consultants and full stack developer Brandon Feldhaus to focus on the VA ODC and AWS resources.
  - Blake Steiner and Yin-Hsuen Chen tasked geospatial and hub developers

ESTO Earth Science Technology Office

7

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

8

Technical Advances: Data Cube Content

- Data Cube Repository and Analysis
  - IoT
    - Flood sensor integration and APIs
  - Flood models
    - TideWatch, DELFT, NWM, STOFS and STACs
  - EO data
    - NASA STAC EO Earthdata
    - Water extents (Landsat, Sentinel 2 OPERA DSWx
    - VLM from Shirzaei, JPL OPERA, and future SWOT and NISAR products
  - ~12 Existing ODC catalog datasets
  - Hub connected Geodata
    - ODU HR Planning District Commission, HR Sanitation District, municipalities, Va. Dept. Emergency Mgt, and HR Biomedical Research Consortium
  - Others
    - NOAA and USGS CoNED, NHD, and Lidar

ESTO Earth Science Technology Office

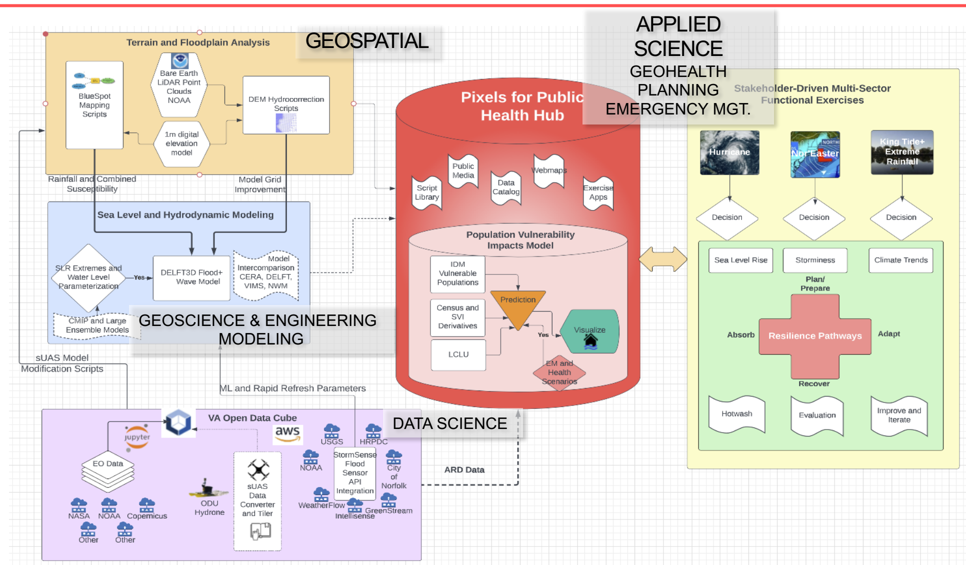
9

Hampton Roads' Coastal DT Twin Framework

Scope	A digital twin to integrate sea-level change, coastal hydrodynamics, and evolving climate and vertical land motion (subsidence) drivers of increased flooding.
Capabilities	• Sea-level and flood forecasting, vertical land motion, satellite and in situ IoT flood observations and hydrologic processes, land use/land cover, and socio-economic activity
Digital Replica	• Digital replica of current state of systems to observe water level and extents; predict flooding; provide information for coastal resilience (adaptation, mitigation, and planning)
Forecasting/Prediction	• Identify future states of tidal, storm, and combined flood events, and how they interact with human systems (disasters, adaptation, and mitigation); Use the DT to test and refine optimized planning for adaptation and hazard mitigation (with EEJ consideration).
Impact Assessment (What-if?)	• How might a city adapt to increasing extent, frequency, and magnitude of flooding? • What-if new information on subsidence can reveal relative risk in flooding? • What-if natural and nature-based features can sustain ecosystems and alleviate floods?
Earth Systems	Ocean (sea-level, water quality, tides), atmosphere (storms and extreme rainfall), and land (VLM , LULC, and shoreline change)
Human Systems	Water and transportation infrastructure, emergency services, and port facilities
Resources	Satellite Earth observations, sea-level and flood models (TideWatch, STOFs, DELFT3D, NWS), IoT flood sensors, building parcel land use fabric, future

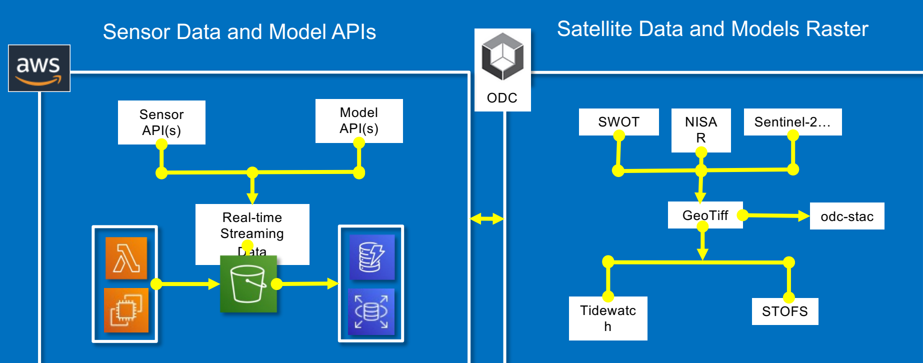
ESTO Earth Science Technology Office

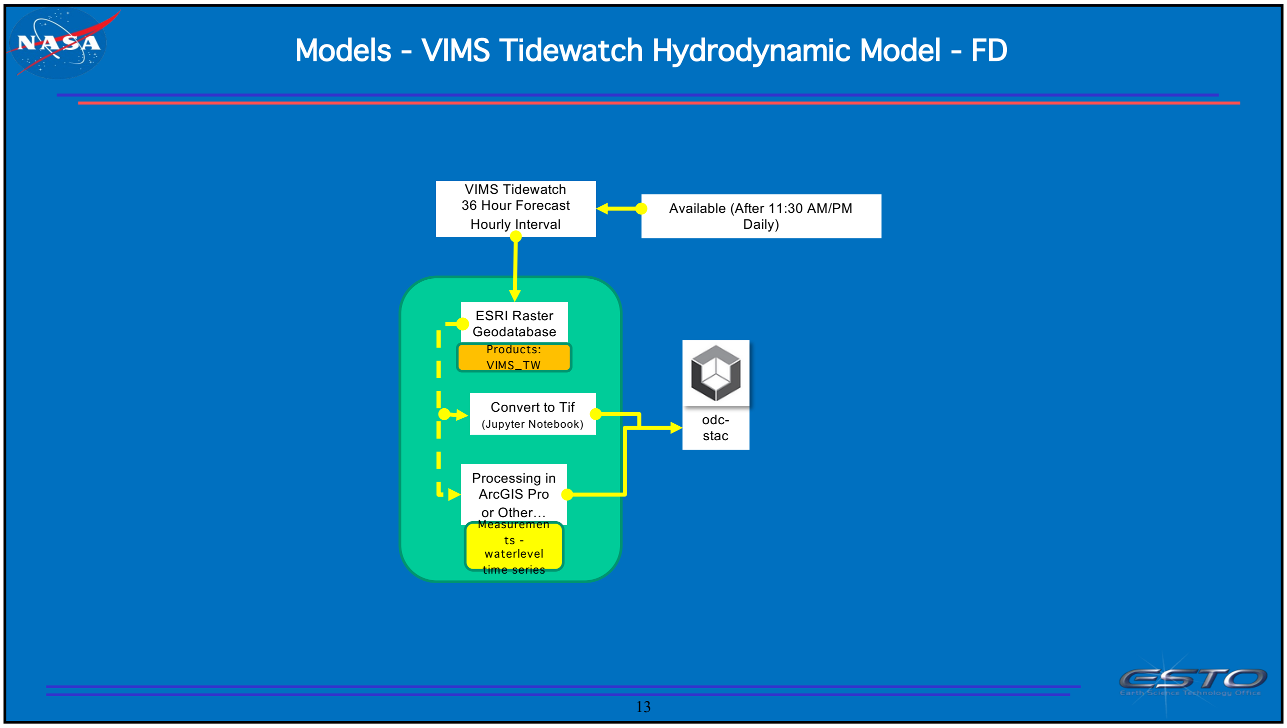
Current High-Level Wireframe

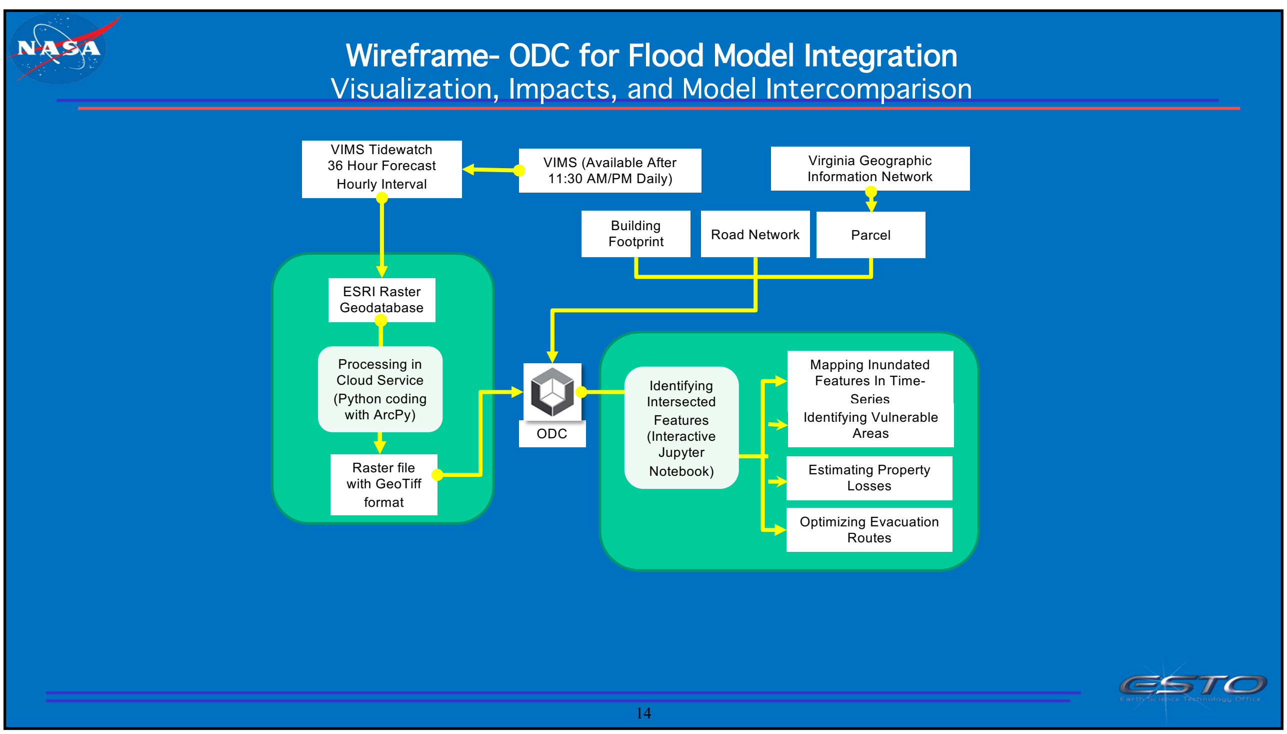


Team editable using Lucidchart

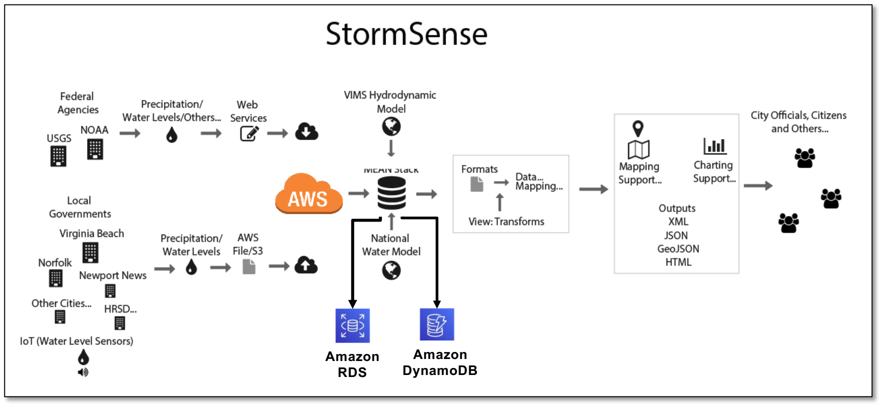
Digital Twin Conceptual Framework for ODC




Models - VIMS Tidewatch Hydrodynamic Model - FD
VIMS Tidewatch
36 Hour Forecast
Hourly Interval
Available (After 11:30 AM/PM Daily)
ESRI Raster Geodatabase
Products: VIMS_TW
Convert to Tif (Jupyter Notebook)
Processing in ArcGIS Pro or Other…
Measurements - waterlevel time series
odc-stac
13


Wireframe- ODC for Flood Model Integration
Visualization, Impacts, and Model Intercomparison
VIMS Tidewatch
36 Hour Forecast
Hourly Interval
VIMS (Available After 11:30 AM/PM Daily)
Virginia Geographic Information Network
Building Footprint
Road Network
Parcel
ESRI Raster Geodatabase
Processing in Cloud Service (Python coding with ArcPy)
Raster file with GeoTiff format
ODC
Identifying Intersected Features (Interactive Jupyter Notebook)
Mapping Inundated Features In Time-Series
Identifying Vulnerable Areas
Estimating Property Losses
Optimizing Evacuation Routes
14

StormSense Schematic



Information Technology | 2/1/2022 | StormSense App Migration Workload

NASA Assets Added to the Data Cube

JPL Opera

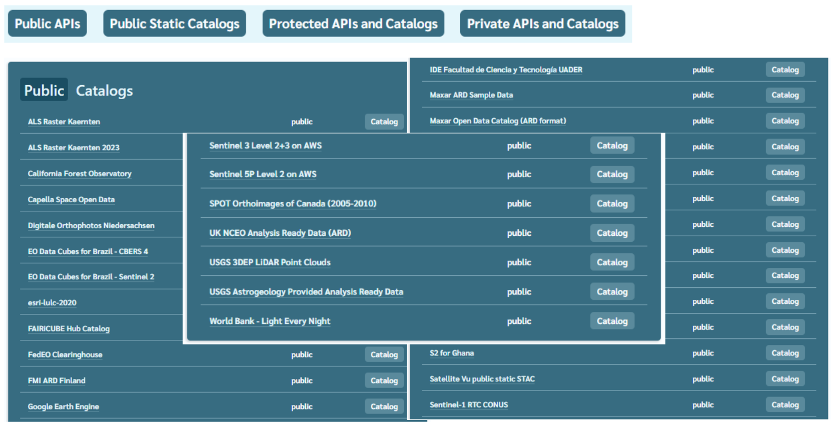
RELEASE	PRODUCTS	PRODUCT RELEASE
1	DSWx-HLS DIST-HLS	April 2023 (provisional)
2	RTC-S1 CSLC-S1	
3	DSWx-S1 DISP-S1	
4	DSWx-NISAR DSWx-SWOT	
5	CSLC-NISAR DISP-NISAR	July 2025

HLS = Harmonized Landsat Sentinel-2; S1 = Sentinel-1; NISAR = NASA-ISRO Synthetic Aperture Radar; SWOT = Surface Water and Ocean Topography.

SWOT

Datasets	Initial Pre-Validated Product Release[1]	Initial Validated Product Release[2]
Level 1 (Nadir Altimeter, Radiometer, GPS, DORIS)	No sooner than July 2023 (Launch +7 months)	Expected in December 2023 (Launch +12 months)
Level 1 (KaRIn)	No sooner than October 2023 (Launch +10 months)	Expected in April 2024 (Launch +16 months)
Level 2	No sooner than October 2023 (Launch +10 months)	Expected in April 2024 (Launch +16 months)

SpatioTemporal Asset Catalogs (STAC)



Data Cube as a Regional STAC Development

- **NOAA 3-D Surge and Tide Operational Forecast System for the Atlantic Basin (STOFS-3D-Atlantic) - STOFS**
  https://aws.amazon.com/marketplace/pp/prodview-jfq56eu763w74
  https://ocean.weather.gov/Loops/ocean_guidance.php?model=STOFS&area=MidAtlanic&plot=cwl&day=1&loop=1

- **NOAA's National Water Model (NWM)**
  https://aws.amazon.com/marketplace/pp/prodview-73iwu7dcfuge2?sr=0-1&ref_=beagle&applicationId=AWSMPContessa
  https://planetarycomputer.microsoft.com/dataset/storage/noaa-nwm
  https://aws.amazon.com/marketplace/pp/prodview-g6lcchc7brshw?sr=0-2&ref_=beagle&applicationId=AWSMPContessa

- **VIMS Tidewatch**
  https://cmap2.vims.edu/SCHISM/TidewatchViewer.html

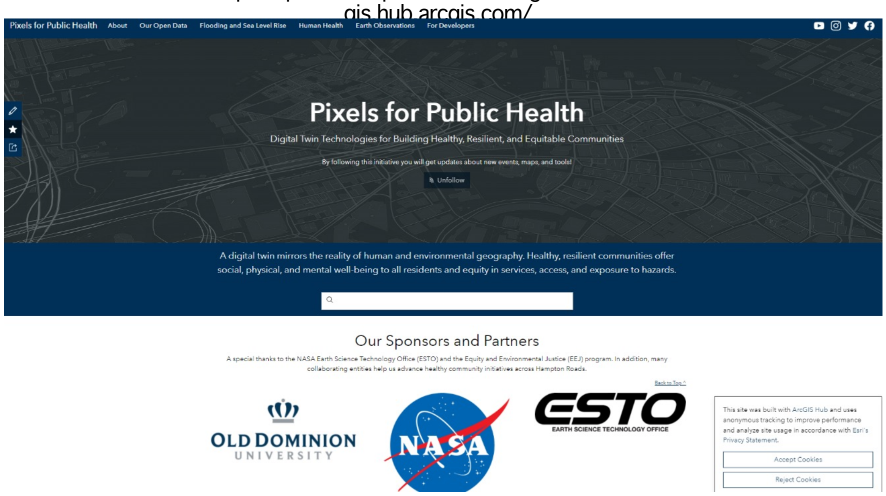
- **Data Cube Open Access and Shared Notebook Repository**
  https://cmap2.vims.edu/SCHISM/TidewatchViewer.html

Geospatial Hub

ArcGIS Hub Technology for Geospatial Specialist and End-Users

tinyurl.com/pixels4health

https://pixels-for-public-health-digital-twin-odu-gis.hub.arcgis.com/



ESTO Earth Science Technology Office

19

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

20

Summary of Accomplishments and Future Plans

- ***Current state: ODC emerging as central to the coastal DT***
  - High readiness to grow content
  - Analytical notebooks current and in development
  - Capacity for expanded EO data (flood model archive and UAS flood extents)
  - Identifying scalability and performance considerations and options
  - Planning derivative documentation, training and publication
  - Hub platform growing to host scores more regional datasets and IO with the ODC
  - Preparing for a visualization front-end for both ODC and Hub (focused on scenarios)
  - Planning a user community workshop to refine scope for prototype demonstration (e.g., EM-styled tabletop exercise, nor'easter, tidal flood, or hurricane surge with combined flooding; public health impact and neighborhood accessibility and EEJ impact dashboard)

- ***Future Critical Tasks***
  - Key identification of data cube performance, scaling and analytical optimization
  - Push on fuller geodata and demographic EEJ community data
  - Sustain the IoT flood sensor network and connectivity (VIMS, VDEM, NOAA-USGS) and model a successful use case (target: King Tide 1Oct2023)
  - Enhance combined flood characterization and model intercomparison tools
  - Assess and support DT documentation and tools (Git, documentation and training, student and scientist onboard; e.g., ODU School of Data Science and DT-relatd conferences, symposia, and publications)

21

Current State – DataCube Catalog Additions

Current Repository

Data	Description
io_lulc	Impact Observatory (ESRI) Landcover Classification
landsat_8_c2_l2	USGS Landsat 8 Collection 2 Level-2 Surface Reflectance
ls5_c2l2_sr	USGS Landsat t Collection 2 Level-2 Surface Reflectance
ls7_c2l2_sr	USGS Landsat 7 Collection 2 Level-2 Surface Reflectance
ls8_c2l2_sr	USGS Landsat 8 Collection 2 Level-2 Surface Reflectance
ls9_c2l2_sr	USGS Landsat 9 Collection 2 Level-2 Surface Reflectance
nasadem	NASADEM from Microsoft's Planetary Computer
s2_l2a	Sentinel-2a and Sentinel-2b imagery, processed to Level 2A (Surface Reflectance) and converted to Cloud Optimized GeoTIFFs

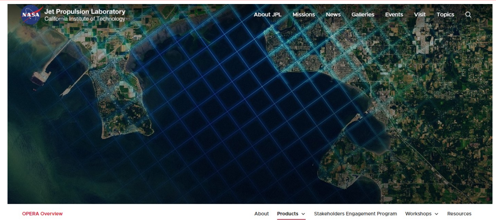
Prioritized Additions

- Sentinel 1 SAR for flood extents
- ODU HydroDem
- Models: NWM, TideWatch, STOFS, DELFT3D flood models
- IOT flood sensors with Heavy.ai
- EPA ESI, ICLUS, Population
- Historic flood extents
- UAV orthoimagery for flood events
- Planet smallsat and Capella Space SAR (Govt. data buy)
- OPERA, SWOT, and NISAR

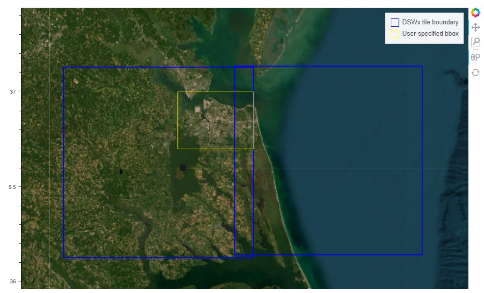
22

Data Cube Example: OPERA Dynamic Surface Water Extent DSWx

- This product is part of *Observational Products for End-Users from Remote Sensing Analysis* (OPERA) Project
- Dynamic Surface Water Extent (DSWx)
- DSWx product is generated from Sentinel-1, NISAR, SWOT, and the Harmonized Landsat and Sentinel-2 data.
- These DSWx products are provided as an analysis ready dataset but are not harmonized across the different sensor.
- More detail: https://www.jpl.nasa.gov/go/opera/products/dswx-product-suite



Study area to filter available DSWx



DataCube Jupyter Notebook to Analyze OPERA DSWx Flood Extent

Generate Flood Maps without downloading OPERA DSWx-HLS products locally

This tutorial demonstrates how to query and work with the OPERA DSWx-HLS Provisional Data products from the cloud (OPERA_L3_DSWX-HLS_PROVISIONAL_V0).

Data Used in the Example:

New collection can be called from PROVISIONNAL_V1

- **30 meter (m) global OPERA Dynamic Surface Water Extent from Harmonized Landsat Sentinel-2 provisional product (Version 0)** - OPERA_L3_DSWX-HLS_PROVISIONAL_V0
  - This dataset contains Level-3 Dynamic OPERA provisional surface water extent product version 0. The data are provisional non-validated surface water extent observations over selective locations and times spanning from February 2019 to September 2022. The input dataset for generating each product is the Harmonized Landsat-8 and Sentinel-2A/B (HLS) product version 2.0. HLS products provide surface reflectance (SR) data from the Operational Land Imager (OLI) aboard the Landsat 8 satellite and the MultiSpectral Instrument (MSI) aboard the Sentinel-2A/B satellite. The surface water extent products are distributed over projected map coordinates using the Universal Transverse Mercator (UTM) projection. Each UTM tile covers an area of 109.8 km × 109.8 km. This area is divided into 3,660 rows and 3,660 columns at 30-m pixel spacing. Each product is distributed as a set of 10 GeoTIFF (Geographic Tagged Image File Format) files including water classification, associated confidence, land cover classification, terrain shadow layer, cloud/cloud-shadow classification, Digital elevation model (DEM), and Diagnostic layer.
  - **Science Dataset (SDS) layers: Pakistan Floods**
    - In 2022, Pakistan's monsoon season produced significant rainfall, devastating floods and landslides, affecting all four of the country's provinces and ~14% of its population [CDP]. Here, we demonstrate how DSWx-HLS can be used to map inundation extent as a result of the monsoon event in Sep 2022.
    - B02_BWTR (Binary Water Layer)
    - B03_CONF (Confidence Layer)

Please refer to the OPERA Product Specification Document for details about the DSWx-HLS product.

Before Starting this Tutorial

A NASA Earthdata Login account is required to download the data used in this tutorial. You can create an account at the link provided.

GitHub link: https://github.com/OPERA-Cal-Val/OPERA_Applications/blob/main/DSWx/Flood/DSWx_FloodProduct.ipynb

OPERA DSWx Example in the Virginia DataCube

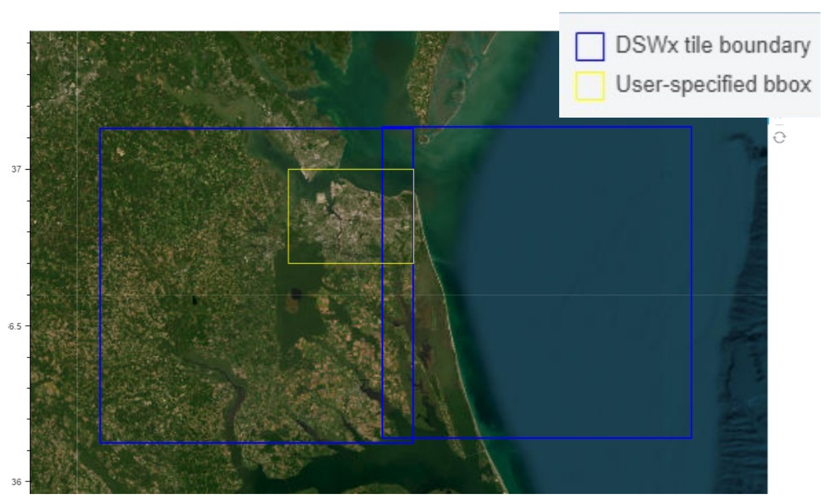
- **10 layers are provided in DSWx product suit**
- **For example, the layer 2 provide binary water layer**

Layer: 2	**Binary water (BWTR)**
Type: UInt8	**Shape (x, y): (3660 × 3660)**
Description: The binary water map is derived from the WTR layer as a union of water classes (open water and partial surface water) into a binary map indicating areas with and without water. This layer is meant to provide users with a quick view for water/no-water. Invalid data classes (snow/ice, cloud/cloud shadow along with adjacent to cloud/cloud shadow, ocean masked, and fill value) are also provided to indicate areas in which the binary classification does not provide water/no-water classification. **Layer classes:** 0: Not Water – an area with valid reflectance data that is not water (class 1) and not snow/ice (class 252), cloud/cloud shadow (class 253), or ocean masked (class 254). 1: Water – an area classified as "open water" or "partial surface water" (see WTR layer). 252: Snow/Ice - an area identified as snow/ice according to input HLS Fmask quality assurance (QA) data. 253: Cloud/Cloud Shadow – an area identified as cloud or cloud shadow or adjacent to cloud/cloud shadow according to input HLS Fmask quality assurance (QA) data. 254: Ocean Masked - an area identified as ocean using a shoreline database with an added margin 255: Fill value (no data).	

Product document: https://d2pn8kiwq2w21t.cloudfront.net/documents/ProductSpec_DSWX_URS309746.pdf

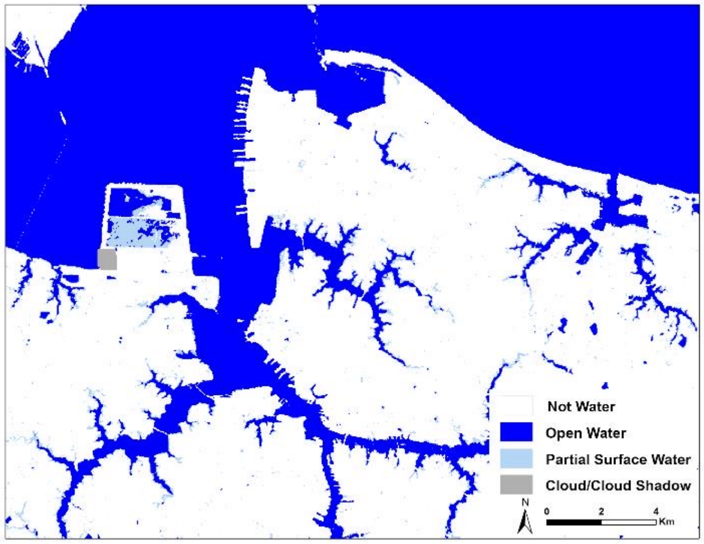
25

DT Study Area to Filter Available DSWx Products



26

OPERA DSWx May 2023 Hampton Roads

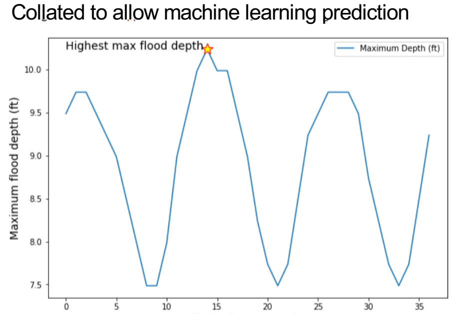


Example 2: VIMS TideWatch inundation model is read daily to map and analyze flood impacts.

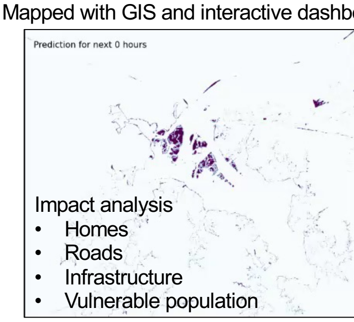
Raw time series hydrology data

	Prediction hour	Minimum Depth (ft)	Mean Depth (ft)	Maximum Depth (ft)
0	0	0.000108	0.816269	9.485605
1	1	0.000065	0.897950	9.735605
2	2	0.000065	0.894644	9.735605

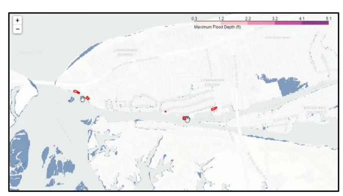
Collated to allow machine learning prediction



Mapped with GIS and interactive dashboards

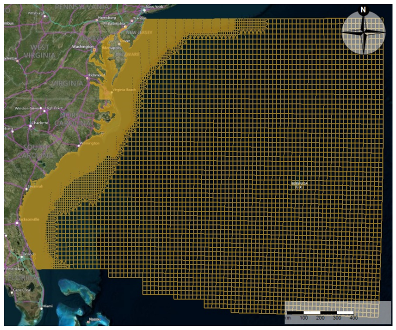


Impact analysis
- Homes
- Roads
- Infrastructure
- Vulnerable population

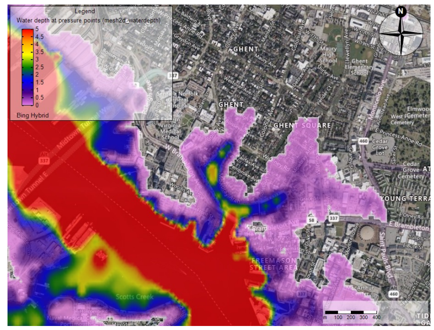


Current State: Hydrodynamic Modeling with DELFT

Model Domain with Grids



Flooding in Hauge (Hurricane Irene 2011)



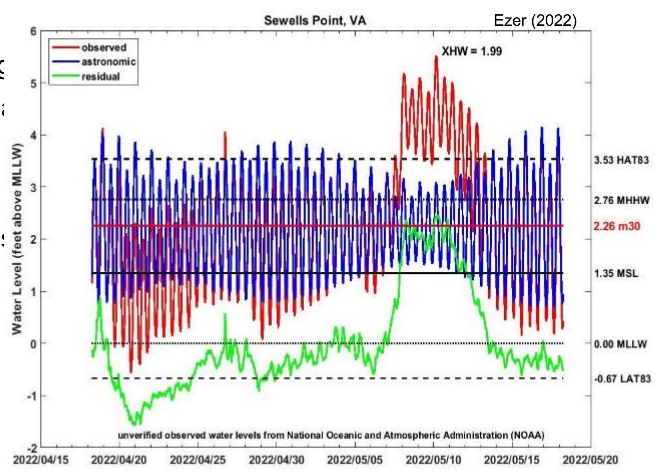
- Extended model grid to water level sensor locations
- Calibrated and validated the model for Hurricane Irene (2011)
- Set up spatiotemporal precipitation input for simulations of compound flooding from rainfall and storm surge

ESTO Earth Science Technology Office

29

Flooding as a Compound Hazard

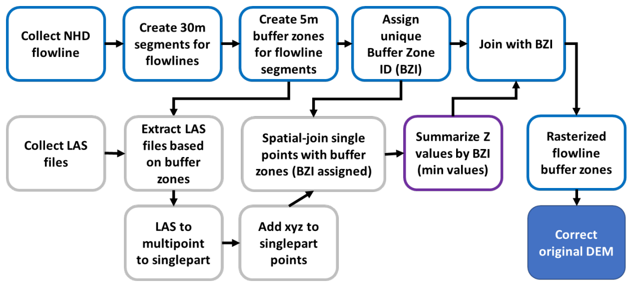
- Two or more flooding sources occur simultaneously or in a short period of time
- Tides
- Wind Tides
- King Tides
- Coastal Storm Surges
- Pre-Cursor Rainfall Event
- Sea-Level Rise
- Subsidence

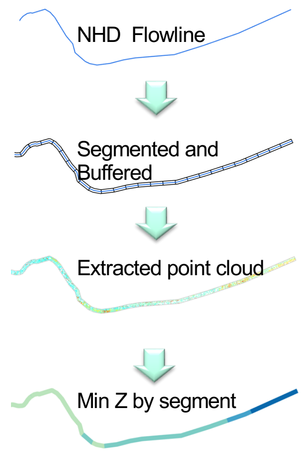


ESTO Earth Science Technology Office

30

Hydro-Correction Model Completed

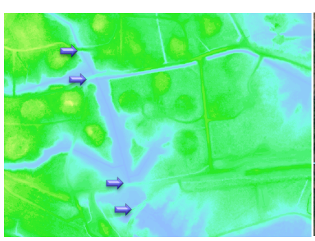




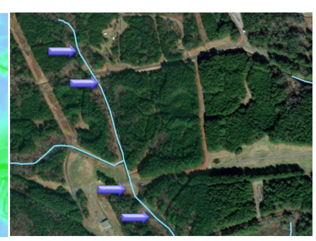
Results of Model Improvements and Flood Prediction Error

With a Digital Twin, we can experiment and assess the uncertainties and impacts of decisions, rather than await an actual event (validation.)

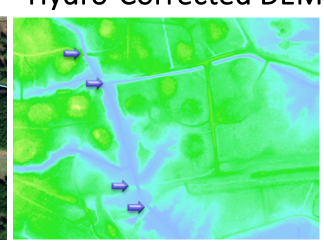
Original DEM



Real World

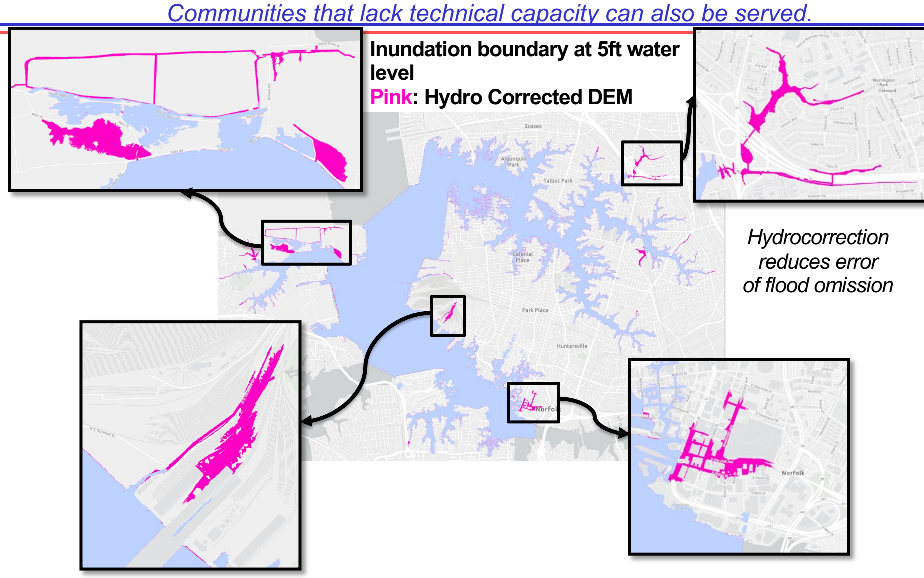


Hydro-Corrected DEM



Resulting Reduction in Flood Error of Omission

Even QL2 Lidar will be improved by this automation.
Communities that lack technical capacity can also be served.

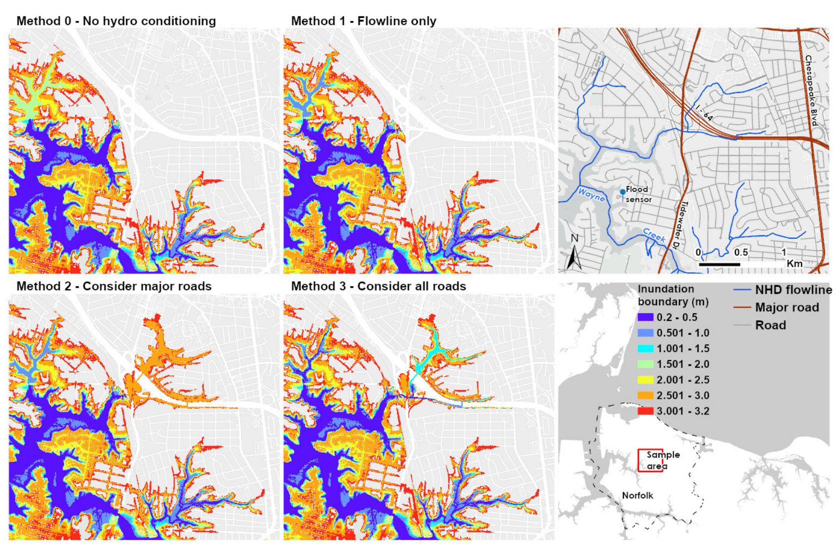


ESTO Earth Science Technology Office

33

Hydro-Correction Sensitivity Results

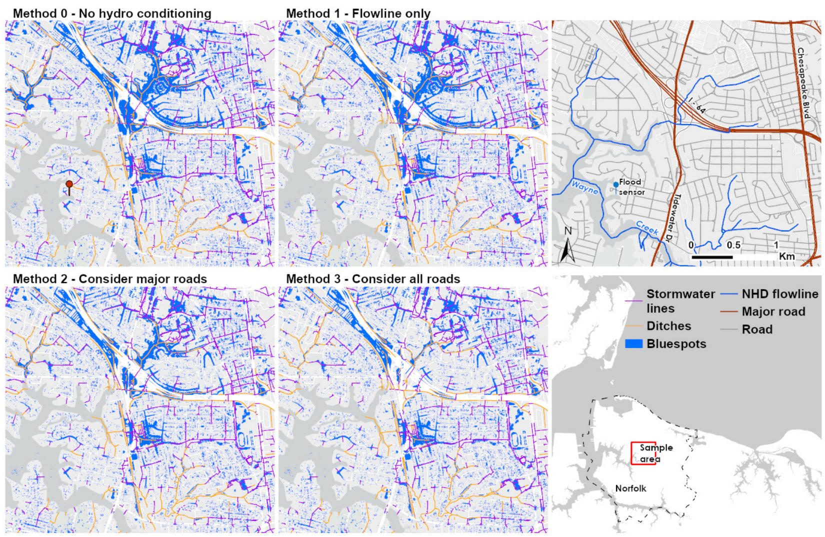
Portable and Allows Fully Automated or Manual Correction



ESTO Earth Science Technology Office

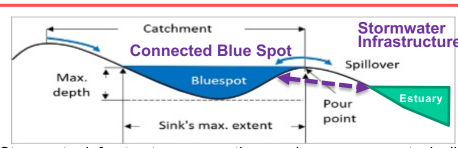
34

Transportation and Flooding

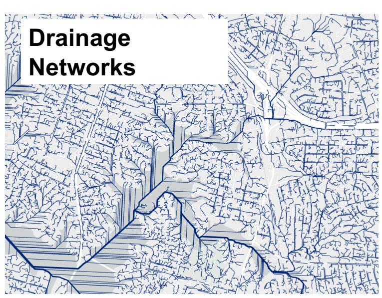


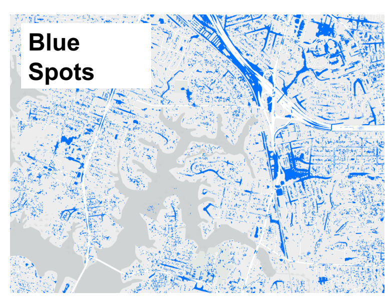
Blue Spots: Landscape Sinks Filled by Rainfall

Landscape sinks, pits, or **Blue Spots** are not necessarily DEM artifacts to be filtered. Rather, they may be important natural or nature-based features, sources and/or conduits for pollutants.



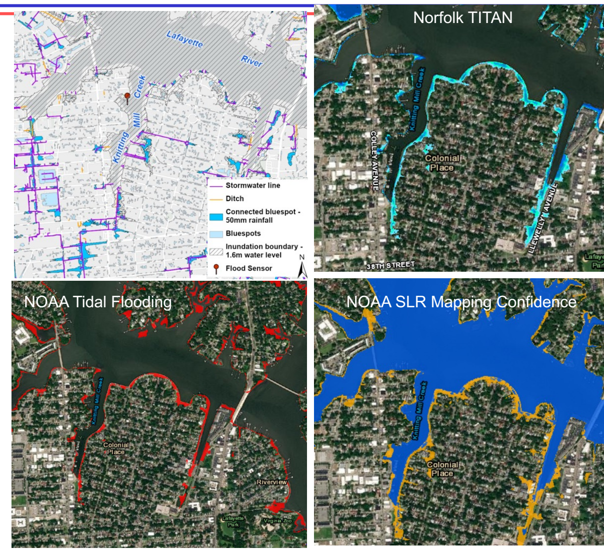
Stormwater infrastructure connections and processes are typically not represented in watershed and sea-level risk assessments (e.g., tidal backflow)





Comparison of Blue Spot Analysis versus Uncertain SLR and Tidal Inundation

- Comparison of hydro-correctd DEM *BlueSpots* to 1.6 m water level with a 50mm rainfall event
- Enhances connections
- Norfolk TITAN
- NOAA SLR Viewer
  - High Tide Flooding
  - Mapping Confidence
- Next, determine the significance for community EEJ and human health impact



37

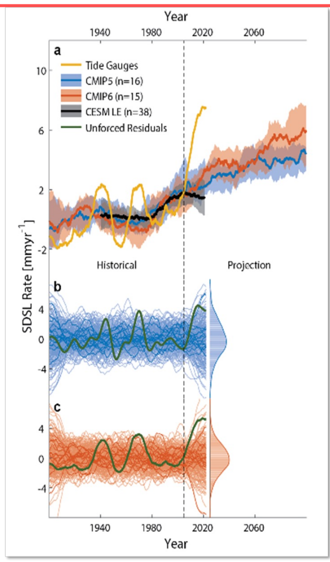
Plan Forward: Hydrodynamic Modeling

- Simulations of compound flooding from rainfall and storm surge in final parameterization phase
- Finalizing workflow for grid edits (hydro-corrections and bathymetry)
- Incorporating sea level rise scenarios in next phase
- Initiated parameter updates protocols
  - IOT-StormSense-datacube ⬄ ODU cluster DELFT3D
  - SLRise forcing
  - Scenarios for Emergency Mgt., NWS, and Health sector next

38

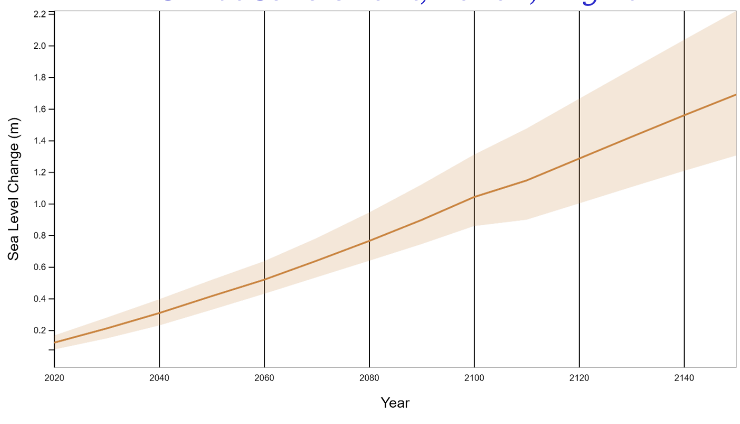
Sea-Level Science Advances

- Accelerating sea-level along the U.S. East Coast has been detected
- Large uncertainty projections owing to future emissions and ice-sheet sensitivity must be reduced.
- We identify accelerations to distinguish between forced sea level and natural variability in Norfolk.
- We approach this using CMIP5, CMIP6, and large ensembles.
- These results will parameterize the flood modeling for future scenarios and prototype demo (yr 3)



Sea Level Rise Attribution and Uncertainty Analysis

SLR at Sewells Point, Norfolk, Virginia



A medium-high scenario of no new major climate policy will realize 1m of relative sea level in Hampton Roads likely before 2100

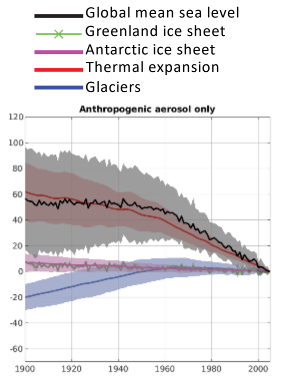
SSP3-7.0 scenario with 1994-2014 baseline

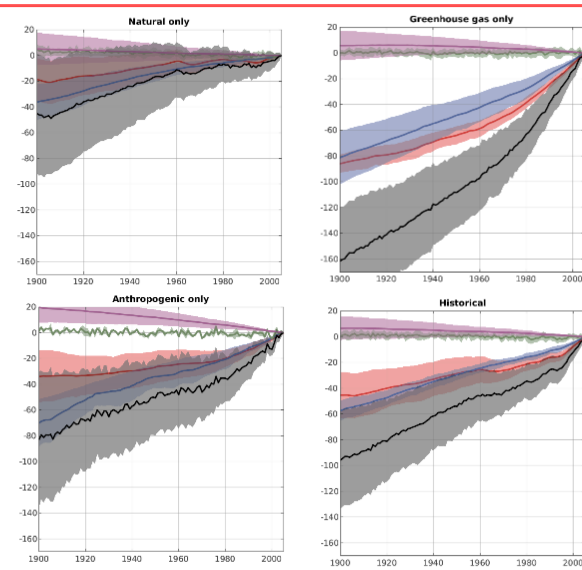
NASA Sea Level Projection Tool

https://sealevel.nasa.gov/ipcc-ar6-sea-level-projection-tool?psmsl_id=299&data_layer=scenario

More detail on SLR attribution and scenarios

- Global mean sea level components in CMIP5 and their attribution to forcings, 1900–2005





ESTO Earth Science Technology Office

41

Current State- GeoHealth

- Prioritizing datacube and GIS data for the hub
- Finalizing stakeholders for scenario protytpe development
- Acquiring inputs for population and flood exposure impact models
  - NASA-USGS-NOAA LULC datasets
  - VA parcels and building footprint planimetrics
  - HRPDC and HRTPO transportation GIS and plans
- Evaluating future socio-demographic scenario data
  - EPA ICLUS
  - USGS ForeSite
  - FUTURES Urban Growth Model

ESTO Earth Science Technology Office

42

GeoHealth Hub Data

- Added *LIFE Project*: Location Intelligence for Food Equity on health, wellness and food accessibility sources and needs.
- Added hydro-corrected DEMs for Norfolk (finishing AOI-wide models)
- Added other priority static layers
  - Underground Storage Tanks and facilities (USTs)
  - Environmental Sensitivity Index (ESI) layers
  - USDA, Census, and CDC social vulnerability indices
  - Air Quality Sensors

ESTO Earth Science Technology Office

43

Ongoing Hub and DataCube GeoData

- Inventory and prioritized 88 geospatial datasets
- Refine detailed wireframes after stakeholder reviews
- Moving to Tier 2 data July-August with HRBRC

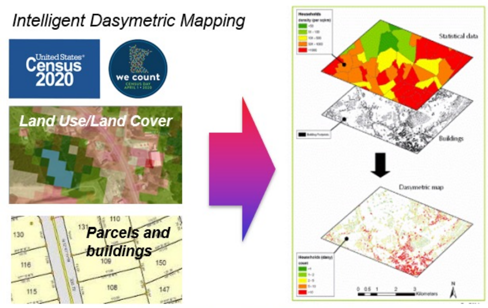
Category	AIST-Priority	Name	Address / Location	Level WRT Inidividual	Level WRT GIS	Sepcial Restrictions
Environment	1	LIDAR NOAA Data (hydroconditioned)	https://coast.noaa.gov/dataviewer/#/	Distribution / Amount of Greenspace within an A	Lat, Lon Grid Box	None
Environment	1	FEMA Flood Zones	https://msc.fema.gov/portal/home	Flood Zones	Lat, Lon Grid Box	None
Environment	1	EPA Enviro Atlas	https://www.epa.gov/enviroatlas	wide array of source threats and social exposure	GIS, web	None
Health	1	Food Desert Locations - Location of Grocery Stores, Restaurants, Delivery, C	City of Norfolk Internal Data	Location of Businesses	GIS - Shape File	????
Health	1	HRBRC LIFE Project food assets, needs, GIS datasets	Chris Davis/ODU AGOL org			
Social	1	Census 2020 in Esri Living Atlas, multi layer	https://www.arcgis.com/home/item.html?id=30d2e10d4d694b3eb4dc4d2e58dbb5a5			
Social	1	Public Housing	City of Norfolk, Census Data	GIS - Shapefile	GIS - shape file	None
Social	1	CDC Social Vulnerability Index (SVI)	CDC	GIS, Shapefile, Web, and Esri Living Atlas	GIS	None
Social	1	"Cutoff" areas / Walkability (TO BE created)	City ????	GIS - Shapefile	????	Spatial Analysis Likely
Transport	1	GIS Network Data for Region AOI				
Environment/Healtl	2	VDEM evacuation zones/Know Your Zone ABCDE				
Environment	2	Environmental permits/Violations: TRI, Hazardous Water, Air Quality permits,	EPA,VA Dept Env Quality, OSHA, City of Norfolk, Hampton Roa	Location of Entities, modeled discharge region	GIS - point source, shapefile, classified remc	None
Health	2	Distance From Home to Primary Care, ER, Pharmacy (TBD Euclidean and Ne	State Level / State Health Dept / City (may be difficult to determi	Distance per group / individual	(may be difficult to determine aggregation le	????
Health	2	Ambulance/Fire points or service areas and drive-times				
Health	2	2021 Food Insecurity Index Conduent Healthy Communities Institute	http://www.ghrconnects.org/index.php?module=indicato	Index for Census Tract	Census tract	????
Social	2	VA Parcels and bldg footprints (possibly also Microsoft)				
Social	2	ADAPT VA BRIC or other social vulnerability indices				
Social/Economic	2	JusticeMap	http://www.justicemap.org/			
Environment	3	Toxic Waste Levels	EPA / Virginia State Office	Waste Level for a specified location	????	????
Social	3	Virginia Crime Data	https://www.vsp.virginia.gov/Crime_in_Virginia.shtm	Total # of Individuals	County	None
Social	3	Norfolk City Crime Open Data	data.norfolk.gov	Total Police Calls By Type	City Block	None
Social	3	Vacant Housing	Census, City of Norfolk	Location	GIS - shape file	None
Social	3	Homeless Shelters	City of Norfolk, Municipal, Google	Location	GIS - shape file	None
Environment	4	Air Quality	EPA / Virginia State Office	Air Quality for a specific location	????	????
Health		Hosptial Data (Sentara; Bon Secours; CHKD; EVMS; Military Hospital)	????	Individual (Sanitized EMR)	Sanitized Address	Yes
Basemap		Community Features (roads, schools, hospitals, churches, residential areas, z	ESRI, Census, City/State	Location of Entities	GIS - Shapefiles/Layers	None
Basemap		Natural Areas (Parks, Presesrves, Wildlife Areas, etc.)	ESRI, Remote Sensing (Landsat), USGS, park service, etc	Location of Entities	GIS - Shapefiles, Layers, Classified Remote	None
Basemap		Agricultural Areas	ESRI, Remote Sensing (Landsat), USGS, park service, etc	Location of Entities	GIS - Shapefiles, Layers, Classified Remote	None
Economic		Norfolk City Permit Open Data	data.norfolk.gov	Indication of if type of permit was applied for / gra	Address	None
Economic		ERSI Community Spending Data	https://www.esri.com/en-us/arcgis/products/arcgis-comm	Community Spending	Census blocks	ERSI License
Economic		Real Estate Information (Transactions, Propoerty values, Tproperty taxes, etc	City/Local Government, City of Norfolk	Address, GIS - Shapefile	GIS - geocoded Address	None
Economic		City of Norfolk Business Licenses	data.norfolk.gov	Indication of if the business is minority / women c	Lat, Lon	None
Economic		Insurance	State Level / State Health Dept / City (may be difficult to determi	Total / Percentrage of Individuals	(may be difficult to determine aggregation le	????
Economic		Business Activity/Trends	City , State Business Filiings, Tax Data	Licenses, Tax Returns	GIS - Geocoded Address (point and/or shap	None
Economic		Employment/Unemployment	Tax Filings, Unemployment filings	Rates	GIS - shape file	None
Economic		Home ownership	Census, City of Norfolk	Location / Ratio	GIS - shapefile with data tables	None
Environment		Green Space Locations	data.norfolk.gov / GIS	Total / Type of Green Space	GPIN (maps to Lat, Lon, Shape file)	None
Environment		NDVI Greeness	https://www.drought.gov/data-maps-tools/ndvi-greennes	Distribution / Amount of Greenspace within an A	Lat, Lon Grid Box	None
Environment		City of Norfolk Flooding data	data.norfolk.gov	Reported Flooding	GIS - shape file	None
Environment		WAZE - Flooding	WAZE API	Reported Flooding Locations	Lat, Lon	None

ESTO Earth Science Technology Office

44

Plan Forward - GeoHealth

- Implement population exposure and vulnerability model
  - Future demographic and land use data integration
  - Hub and datacube interconnection
  - Codeshare via GIT
  - Community accessibility metrics
- Share hub site with stakeholders and wireframe the scenario and prototype exercise
- Identify outreach and sharing communication plans
  - GeoHealth Community of Practice



45

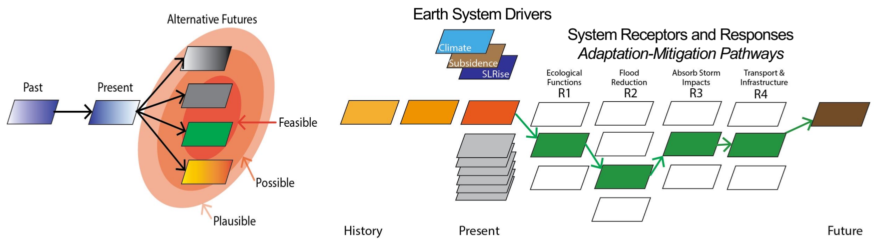
Summary Actual or Potential Infusions

- Infusions:
  - Adoption of ODC and Hub technology by HRSD, HRPDC, and NWS Wakefield, VDEM, Norfolk and other Em. Mgt.
- Knowledge transfer:
  - Sharing with peers and stakeholders (Coastal GeoTools, AGU, AWS, VA and topical conferences and journals)
  - Developing guide and training materials; educational materials for ODU data science, ocean-earth science, geography and coastal engineering and emergency mgt.
- Technology transfer:
  - Public open-source software and utilities via Git and Esri Hub
- Transition:
  - Replication and application to other Digital Twin systems
    - Coastal transportation
    - Public health
    - Ocean energy and ecosystems

46

Enabling What-If Scenarios

We can imagine future states……..and model alternative plans and future conditions to Identify resilience adaptation pathways.



- Future sea-level rise and coastal flooding
  - Impact assessment
  - Mitigation and adaptation
  - Equity and environmental justice

After Steinitz (2012) *A Framework for Geodesign: Changing Geography by Design*

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Actual and Potential Collaborations

- *Actual*
  - NASA, NOAA NESDIS, USGS VA Water Center
  - National Weather Service- Wakefield, VDEM Flood Program
  - Hampton Roads Planning District Commission, HR Sanitation District, HR Biomedical Research Consortium
  - VIMS, EVMS, GeoHealth Community of Practice
  - NISAR Early Adopter Program and NASA SWOT, OPERA, VT VLM
- *Potential*
  - Other coastal-marine Digital Twin initiatives (CNES, MARI Ireland)
  - NOAA CoastWatch Chesapeake Bay and CEOS Ches. Bay
  - DOE Jefferson Natl. Lab
  - Cooperative Institute for Climate and Earth System Studies (CICEES)
  - UNC (RENCI/NC-CERA ADCIRC), Georgia Tech (Savannah), and The Citadel and Coll. of Charleston (IOT flood sensor networks and model domains)

ESTO Earth Science Technology Office

49

Summary

- Full steam ahead modeling, DataCube and hub implementation
- DataCube repositories being stocked.
- Leaning into geohealth, population impact, and scenario modeling next quarter
- SLR scenarios and probabilistic forecasts in preparation for future flood modeling
- Hydrodynamic model-IOT-EO integration is initiating
- Gathering insights and awareness of related DTs, interest in cloud-based registries, and open data sharing and collaboration
- Emerging collaborations within NASA and with CNES, VIMS, and NOAA

ESTO Earth Science Technology Office

50

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Publications

- Allen, Tom & McLeod, George & Richter, Heather & Nielsen, Alex. (2022). Digitally Twinning Coastal Resilience via Multisensor Imagery, in Situ Sensors, and Geospatial Analysis. 4739-4742.*IGARSS IEEE XPLORE* 10.1109/IGARSS46834.2022.9883133.

- Dangendorf, S., Hendricks, N., Sun, Q., Klinck, J., Ezer, T., Frederikse, T., ... & Tornqvist, T. (2022). Acceleration of US Southeast and Gulf Coast Sea-Level Rise Amplified by Internal Climate Variability, under review in *Nature Communications*, preprint online available.

- Allen, T.R., A. Garnand, R. Stuart, G. McLeod, D. Eulie, M.A. Polk. Advancing Estuarine Shoreline Change Analysis using Uncrewed Autonomous Systems. *Estuary Research – Recent Advances,* InTechOpen (in press, accepted 5 June 2023).

- Hutton, Nicole & McLeod, George & Allen, Tom & Davis, Christopher & Garnand, Alexandra & Richter, Heather & Chavan, Prachi & Hoglund, Leslie & Comess, Jill & Herman, Matthew & Martin, Brian & Romero, Cynthia. (2022). Participatory mapping to address neighborhood level data deficiencies for food security assessment in Southeastern Virginia, USA. *International Journal of Health Geographics*. 21. 10.1186/s12942-022-00314-3.
- Upcoming: Frontiers in prep, IGARSS 2023, *Computers & Geosciences* or *IJGIS* in prep.

Thank You and Welcome Questions

Dr. Thomas R. Allen
Professor of Geography
Department of Political Science and Geography
Institute for Coastal Adaptation and Resilience
E-mail: tallen@odu.edu

OLD DOMINION UNIVERSITY

With credit to Co-Investigators and Partners:
Drs. George McLeod, Navid Tahvildari, Yin-Hsuen Chen, Blake Steiner, Heather Richter, Sunghoon Han, and Brandon Feldhaus, ODU
Drs. Sonke Dangendorf and Qiang Sun, Tulane University
Dr. Oguz Yetkin, Brian Terry, and Josh Baptist, Analytical Mechanics Associates
Mr. Sridhar Katragadda, City of Virginia Beach
Mr. Jeff Orrock, National Weather Service, Wakefield, Va.
Mr. Chuck Kirby, Virginia Innovative Partnerships Corporation

ESTO Earth Science Technology Office

53

List of Acronyms

- ASV — Autonomous Surface Vessel
- AWS — Amazon Web Services
- CMIP — Coupled Model Intercomparison Project
- DEM — Digital Elevation Model
- DSWx — Dynamic Surface Water Extent
- DT — Digital Twin
- ESRI — Esri.com
- EVMS — Eastern Virginia Medical School
- HRBRC — Hampton Roads Biomedical Research Consortium
- HRPDC — Hampton Roads Planning District Commission
- HRSD — Hampton Roads Sanitation District
- JPL — Jet Propulsion Laboratory
- NC-CERA — North Carolina Coastal Emergency Risk Atlas
- NHD — National Hydrologic Dataset
- NWM — National Water Model
- NISAR — NASA-ISRO SAR Mission (Synthetic Aperture Radar)
- NOAA — National Oceanographic and Atmospheric Administration
- ODC — Open Data Cube
- OPERA — Observational Products for End-Users from Remote Sensing
- RENCI — Renaissance Computing Institute (UNC)
- SCHISM — Semi-Implicit Cross-Scale Hydroscience Integrated System Model
- STAC — Spatial-Temporal Asset Catalog
- STOFS — Storm Surge & Tide Operational Forecast
- SWOT — Surface Water and Ocean Topography
- UAV — Uncrewed Aerial Vehicle
- UNC — University of North Carolina
- USGS — United States Geological Survey
- VIMS — Virginia Institute of Marine Science

ESTO Earth Science Technology Office

54

Open Data Cube Resources

PostgreSQL 10+
Python 3.8+

STAC
https://stacspec.org/en/

ODC STAC
https://docs.digitalearthafrica.org/en/latest/platform_tools/odc_stac.html#Set-AWS-Configuration

ODC Core Concepts
https://datacube-core.readthedocs.io/en/latest/about-core-concepts/index.html

ODC Core Architecture
https://www.opendatacube.org/_files/ugd/3632b4_269d1d61d7f04677a1d32278042aa51a.pdf

Products and Measurements
https://docs.digitalearthafrica.org/en/latest/sandbox/notebooks/Beginners_guide/02_Products_and_measurements.html

ODC FAQ
https://www.opendatacube.org/faq

DEA Tools Package
https://docs.dea.ga.gov.au/notebooks/Tools/index.html

55

An Analytic Collaborative Framework for the ESO

Arlindo da Silva (PI, NASA/GSFC Code 610.1)
Ben Poulter (Co-I, NASA/GSFC Code 610.1)

AIST-18/21-0107 Annual Technical Review
23 June 2023

(see slide X for Full Team)

Proposal Team

Name	Org	Code	Role	Team Member Responsibility
Arlindo da Silva	GSFC	610.1	PI	Overall project management; Nature Runs and Ob. Systems simulations
Ben Poulter	GSFC	618	Co-I	Coordination of land surface activities; modeling, ob. systems simulations
Alexey Shiklomanov	GSFC	610.1	Co-I	Cyberinfrastructure coordination, land surface science, bio diversity demo
Patricia Castellanos	GSFC	610.1	Co-I	Lead simulations of lidar/polarimeter measurements
Reed Espinosa	GSFC	613	Co-I	Lead GRASP lidar/polarimeter retrievals, retrieval assessment demo
(Melanie Follette-Cook)	GSFC	612	Co-I	Coordination of applied science engagement activities
Fei Liu	GSFC	614		Year 1: tobac for aerosol plumes
David Thompson	JPL	-	Co-I	ISOFIT deployment and SBG retrieval algorithms
Derek Posselt	JPL	-	Co-I	ParOSSE deployment and simulation of radar/radiometer measurements
Weile Wang	ARC	-	Co-I	Development and production of Ames surface reflectance dataset
Bryce Currey	GSFC	618	PostDoc	Land-surface Nature Run generation (LPJ, PROSAIL)
Ravi Govindaraju	GSFC	610.1	Supp	Scripting support, dataset sampling, jobs execution
Garrison Vaughan	GSFC/GST	606.2	Supp	Nature Run generation, deployment/optim. of complex HEC applications (Shared with Clune)
Johana Chazaro Cortes	GSFC/NAV TECA	606.2	Supp	Deployment of the SMCE environment on AWS and its maintenance. Cloud Optimized Dataset Development of Jupyter notebooks for algo. workbench, demo activities

An Analytic Collaborative Framework for the ESO

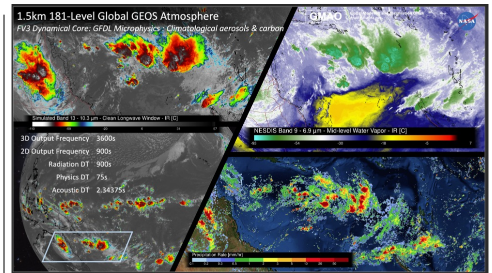
PI: Arlindo da Silva, NASA/GSFC

Objectives

Develop integrated Earth science frameworks that mirror the earth by a proxy digital construct includiung high resolution earth system models and synthetic measurements with an integrated set of analytic tools to enable the start of next generation of science discoveries and evidence-based decision making well before launch.

Provide a science-based, realistic environment for agile development of mission algorithms and data products.

Technologies include global cloud resolving models, synthetic surface reflectances, observation simulators, retrieval systems, and a collaborative cyberinfrastructure enabled by Jupyter notebooks on a cloud environment.



Approach

Consists of 3 interconnected building blocks:

1. Cloud-optimized representative datasets for ESO missions and the PoR to serve as basis for developing and prototyping an Analytic Collaborative Framework
2. An algorithm workbench for enabling experimentation and exploration of single instrument and synergistic algorithms
3. A series of concrete Open Science demonstrations including use cases that span science discovery and end-user applications with direct societal impact.

Co-Is: B. Poulter, A. Shiklomanov, P. Castellanos, R. Espinosa, M. Follette-Cook (MSU), D. Thompson, D. Posselt (JPL), W. Wang (ARC)

Key Milestones

- Project Starts 08/22
- Requirements, architecture, design 12/22
- Nature Run Generation 02/23
- Level B & C simulations 10/23
- Algorithm Workbench prototype 02/24
- Science & App Demos 07/24

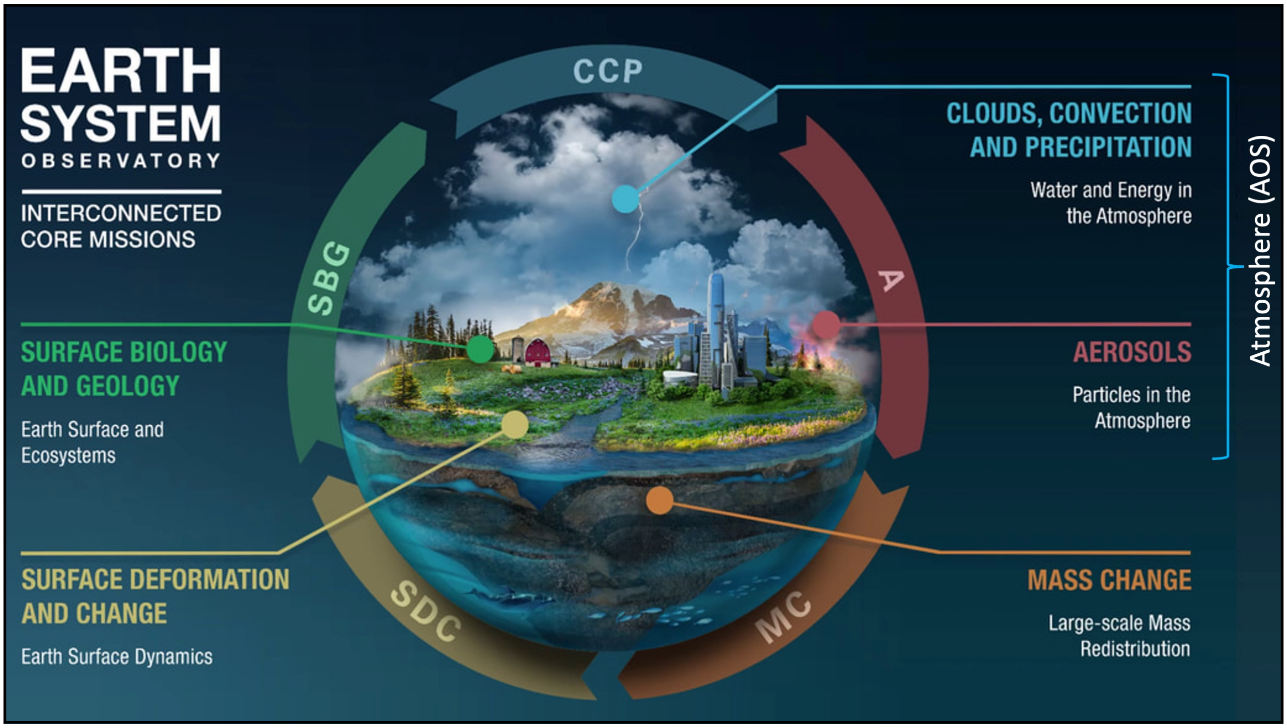
TRL_{in} = 3 **TRL_{curr} = 3**

ESTO Earth Science Technology Office

2023-06-23 3

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms



Project Objectives

Science and Technology Objectives

- UseCloud-optimized representative datasets for the ESO missions and the PoR for prototyping an *Analytic Collaborative Framework*
- An *Algorithm Workbench* for enabling experimentation and exploration of synergistic algorithms involving the AOS, SBG and the PoR
- Science demonstrations: use of synthetic data to demonstrate capability of new observing system
  - SBG atmospheric correction
  - Evaluate lidar-polarimeter retrievals
  - Explore synergistic algorithms

Application Objectives

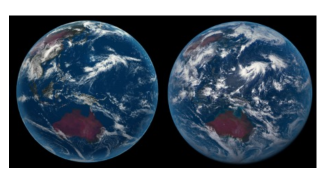
- Concrete Open-source Science demonstrations including use cases that span science discovery and end user applications with direct societal impact.
  - Tracking of PyroCbs
  - Solar energy application
  - Bio-diversity application

Benefits to Earth Science

- Bridges the gap between advanced earth system modeling and realistic simulation of future observing systems
- Early development of new concepts in model-data fusion enabling science investigations aimed at elucidating physical processes and their representation in models.

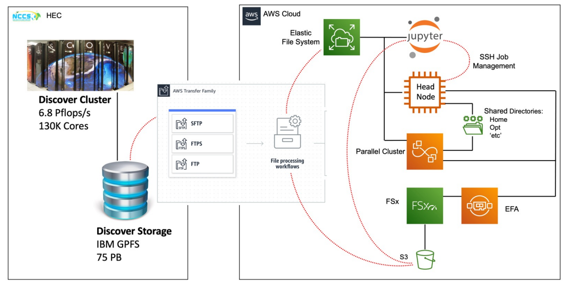
Relationship to ES Digital Twins

- Develops a **Digital Replica** of the future ESO observing system based on realistic, high-resolution Earth system model simulations
- Provide a framework for observing system simulations and tools to conduct *what-if* investigations and trade studies for the benefit of NASA missions in development.

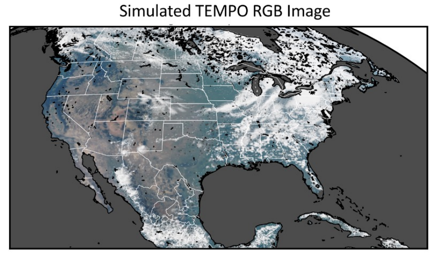


Technology Approach

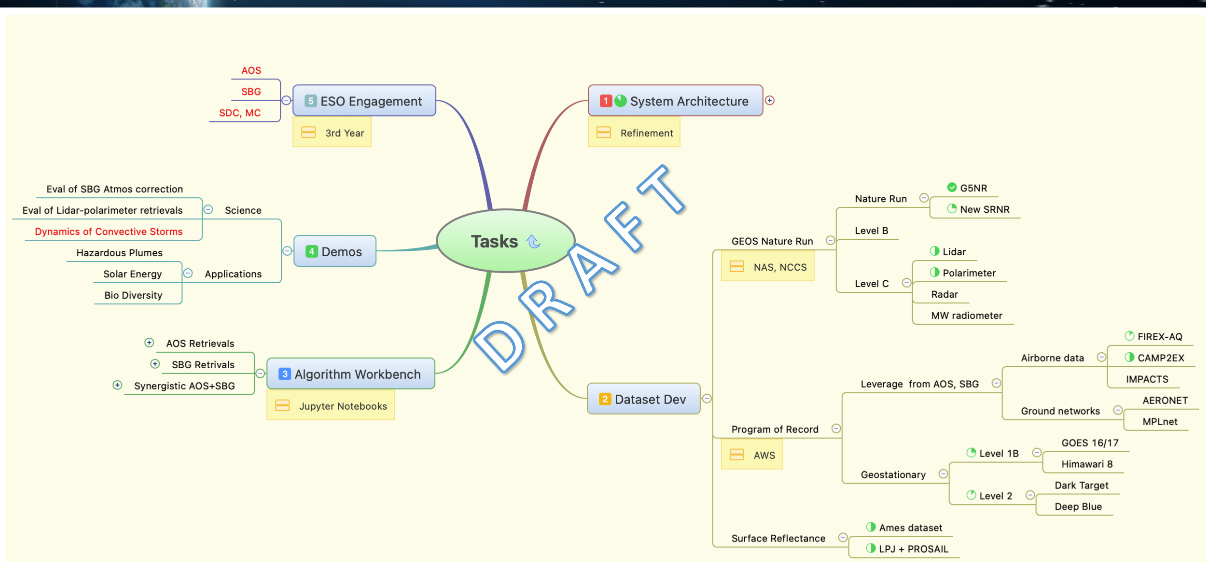
- Based on a global storm resolving GEOS Nature Run, simulate the AOS observing system
- Simulate SBG reflectances at TOA
  - Synthetic surface reflectances
    - Ames MODIS-based spectral filling
    - LPJ/PROSAIL model
  - Storm resolving atmosphere with 20 aerosols
- Implement an Algorithm Workbench on AWS for developing testing and evaluating AOS, SBG and synergistic AOS+SBG retrievals
  - ➢ Uncertainty quantification
- Open-source Science demonstrations by exploring advanced data analysis and easy access to cloud-optimized (simulated) datasets.
  - Explore use of IceSat2 Slide-rule framework



Simulated TEMPO RGB Image



Technology Approach



Presentation Contents

✓ Background and Objectives

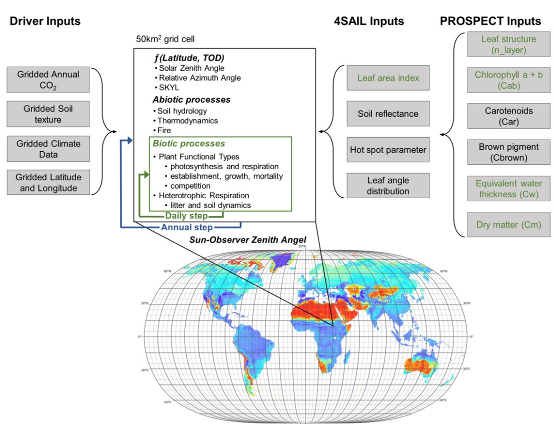
○ Technical and Science Advancements

○ Summary of Accomplishments and Future Plans

○ Actual or Potential Infusions and Collaborations

○ Publications - List of Acronyms

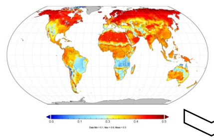
LPJ-PROSAIL



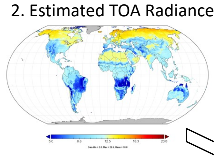
- Simulates global dynamic surface reflectance data for spaceborne imaging spectroscopy missions.
- Dynamic global vegetation model, *LPJ* + canopy radiative transfer model, *PRO4SAIL*.
- Gridded input variables drive process-based vegetation dynamics of 10 plant functional types and downstream biogeochemical and hydrological cycling.
  - Atmospheric Inputs: CRU, MERRA-2 or km-scale Nature Run (in progress)
  - HWSD or USDA global soils
  - CO_2
  - LUH v2 land use change data
- Variables cross-walked between LPJ and PROSAIL (green in the diagram) to estimate globally-gridded, top-of-canopy reflectance data at the same time step as LPJ.
- Reflectance (Level C) data products:
  - Globally-gridded top-of-canopy reflectances for 1980-2021.
  - Daily or monthly time step
  - 0.5° globally gridded
  - 400-2500nm @ 10nm

LPJ-PROSAIL: OSSE Capabilities

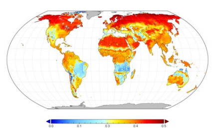
1. Simulated Sfc Reflectance



2. Estimated TOA Radiances

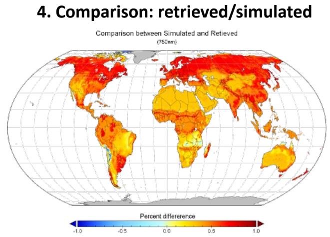


3. Retrieved Sfc Reflectance



- LPJ-PROSAIL linked with ISOFIT (Imaging Spectrometer Optimal Fitting) Optimal Estimation (OE) technique
  - Standard: Uses MODTRAN RTM and an instrument model + atmospheric AOD and H2O data (currently from MERRA2) to estimate TOA radiances (forward mode) and retrieved reflectances (inverse mode).
  - Innovation: use detailed aerosol state from NR and vector RT to calculate TOA radiances
  - Globally gridded TOA radiance data products w/ same specifications as surface reflectance data products (0.5° gridded, 400-2500 nm, 10nm intervals, daily/monthly).
- LPJ-PROSAIL can also produce TOA radiances for each functional type

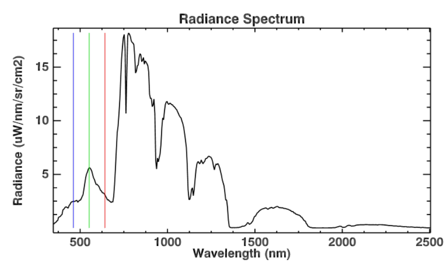
4. Comparison: retrieved/simulated

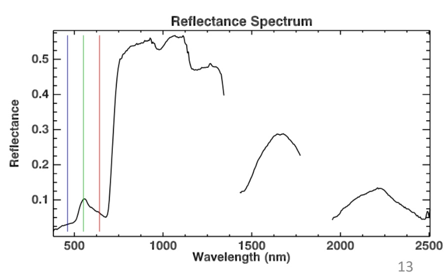


This OSSE capability provides end-to-end modeling support and uncertainty quantification for SBG with known traceability, controlled environment and instrument parameterization.

ISOFIT

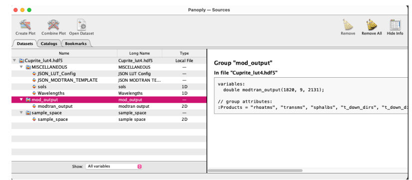
- ISOFIT is a codebase for manipulating and inverting radiative transfer models for VSWIR imaging spectrometers.
- Most commonly, it is used to solve the **atmospheric correction problem** – inverting the radiance at sensor to estimate the surface reflectance and atmospheric state.
- Forward models supported include MODTRAN and neural network emulators
- Inversion approaches include Bayesian Optimal Estimation
- ISOFIT is currently in operational use for the EMIT imaging spectrometer mission and the candidate approach being prototyped for SBG.

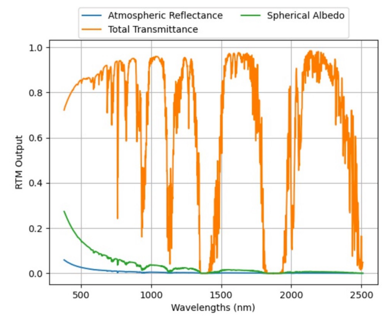




ISOFIT Advancements

- ISOFIT has been integrated into the DISCOVER system
- We have run comprehensive radiative transfer simulations to encompass diverse global atmospheric conditions, and delivered these lookup tables to the team
- We have developed a new hierarchical (HDF5) file format for radiative transfer lookup tables to enable their use in cloud environments, and are testing it with the existing retrieval code. Advantages:
  - Efficiency and Compactness: The new HDF5 file format reduces space usage significantly, achieving at least 50% size reduction compared to the original file format.
  - Hierarchical Structure: Our approach provides a convenient and organized hierarchical structure, housing both the inputs into the RTM and processed outputs, ready for use.
  - Standardization Across RTMs: The HDF5 format is RTM agnostic, enabling seamless integration with multiple Radiative Transfer Models, such as MODTRAN, 6S, and others.
  - Metadata Inclusion: Our format encapsulates comprehensive metadata, including sample space ranges, RTM configuration input, the wavelength grid, and more.
  - Flexibility and Variety: The new format has the versatility to accommodate diverse atmospheric coefficients, including standard and multipart transmittance.

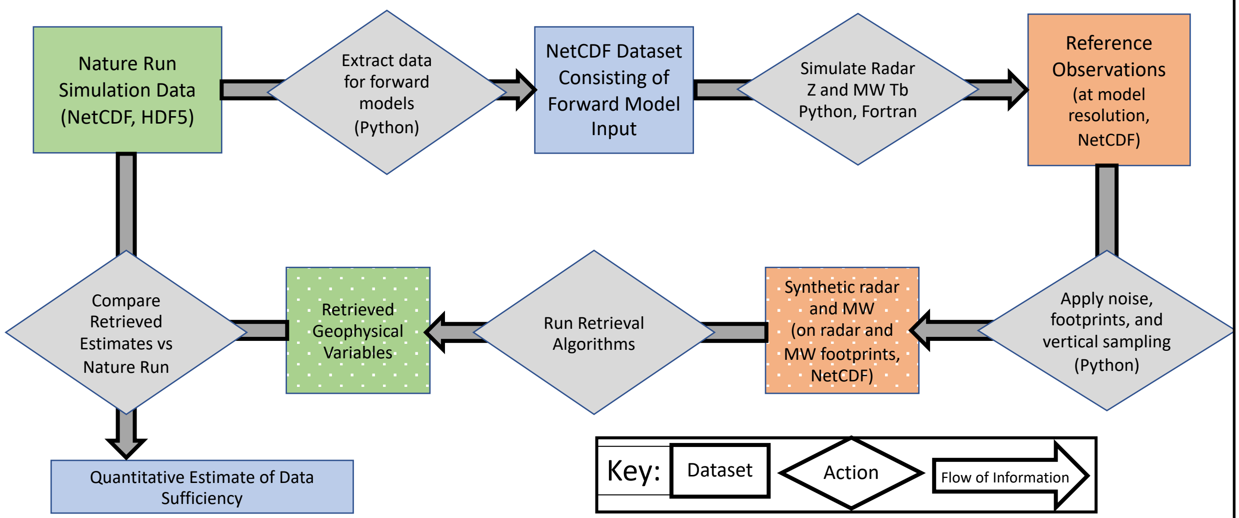




ParOSSE

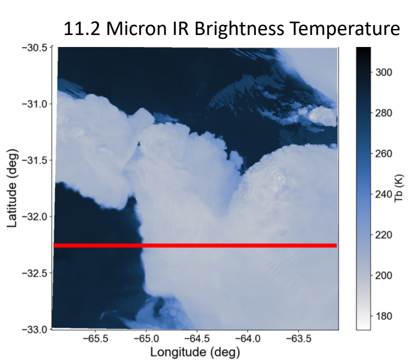
- ParOSSE has 2 components
  - ParMAP flexible parallelism: fast execution on laptop, workstation, cluster, and cloud
  - Experiment curation: managing an end-to-end workflow
- Simulators within ParOSSE have 2 parts
  - Radiative transfer: Convert nature run output to passive or active MW
  - Instrument model: Convert full resolution noise-free dBZ or Tb into noisy and blurry instrument-specific measurements
- Radar and MW simulators
  - Frequency range from 6 to 800 GHz
  - Fast and accurate absorption by gases and hydrometeors, and hydrometeor scattering
  - Well tested on Ku, Ka, and W-band radar, and on 87 – 181 GHz microwave.

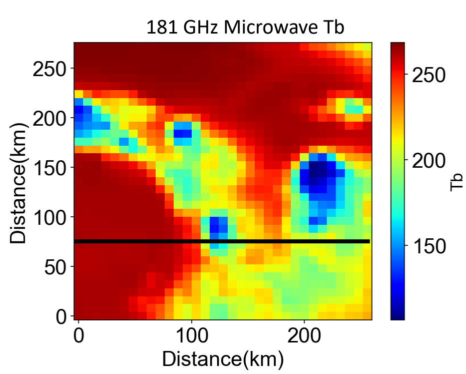
ParOSSE Workflow

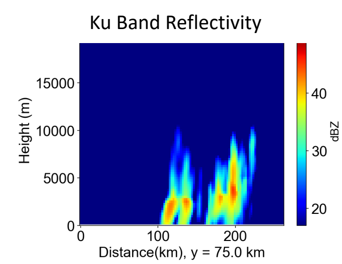


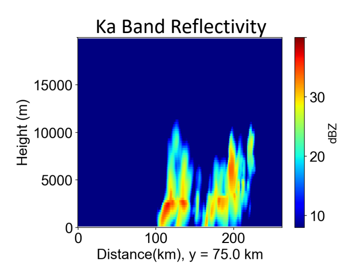
Radar & MW Radiometers Advancements

- Radar and MW forward models can ingest a wide variety of nature run datasets: GEOS, WRF, RAMS, DYAMOND global simulations
- Shown are examples from a high-resolution simulation of convection 14 December 2018, Argentina



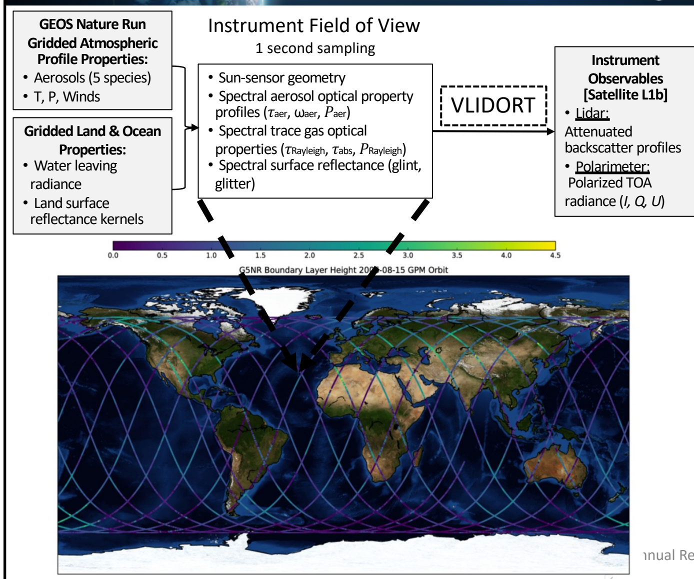






ESTO Earth Science Technology Office

GEOS Instrument Simulator



- Simulates instrument observables for lidars and polarimeters
- VLIDORT radiative transfer model implemented on high performance computing environment
- Gridded atmospheric input variables from GEOS Nature Runs
- Gridded land and ocean properties from MODIS
  - Will be adapted to use LPJ-PRO4SAIL simulations
- Level C data products:
  - Top of the atmosphere spectral *I, Q, & U*
  - Vertical profiles of aerosol attenuated backscatter
  - Global orbits of instrument field of view
  - 1-second time step

ESTO Earth Science Technology Office

Presentation Contents

✓Background and Objectives

✓Technical and Science Advancements

○Summary of Accomplishments and Future Plans

○Actual or Potential Infusions and Collaborations

○Publications - List of Acronyms

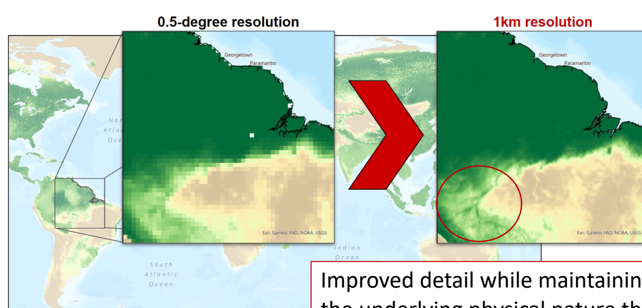
Current State - Highlights

- Refined SW requirements and scope of activities
- Established AWS environment using SMCE DaskHub and pcluster
  - Working with SMCE maintainers tword better integration of SMCE and plcluster environment
- Decided on GitHub and GitHub flow for software configuration management; still working through licensing framework
  - Ideally would like to use same *system* framework as we have adopted for the GEOS Earth System Model
- Started PoR dataset acquisition
  - GOES data from MODAPS, Himawari Data from NOAA S3 buckets
  - Merged airborne data from CAMP2EX for all flights
- Cloud optimized dataset strategy: avoid conversion to zarr, work with existing netCDF
  - Use Kerchunk as performance benchmark for S3 data
  - Exploring strategies for optimized range gets within the NetCDF API: H5Coro as a NetCDF plugin
- Multi-year simulation of LPF/PROSAIL at 50 km resolution completed
- Generation of multi-year Ames surface reflectance completed

LPJ-PROSAIL at km-Scale

Next steps

- Forcing LPJ-PROSAIL with the 1-km scale nature run.
- Challenges:
  - Computational complexity
  - Data output volume >50x
- Enabling Digital Twin capabilities of LPJ via atmosphere-land coupling and land management activities
- See GPP example.

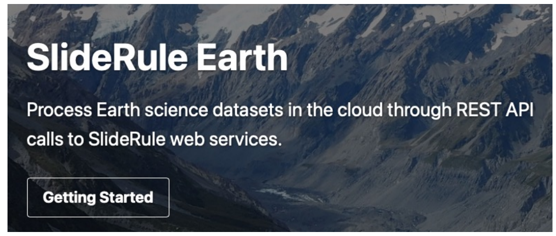


Improved detail while maintaining the underlying physical nature the dynamic global vegetation model.

Slide Rule Earth

SlideRule Earth
science data processing as a service

- Slide Rule Earth enables low-latency access to data products using processing parameters supplied at the time of the request.
- The Slide Rule backend package H5Coro provides cloud optimized access to HDF-5/NetCDF-4 files on S3.
- Investigating extension of the NetCDF C library by means of a H5Coro plug-in.



SlideRule runs in AWS us-west-2 and currently has access to ICESat-2, GEDI, Landsat, and ArcticDEM datasets stored in S3.

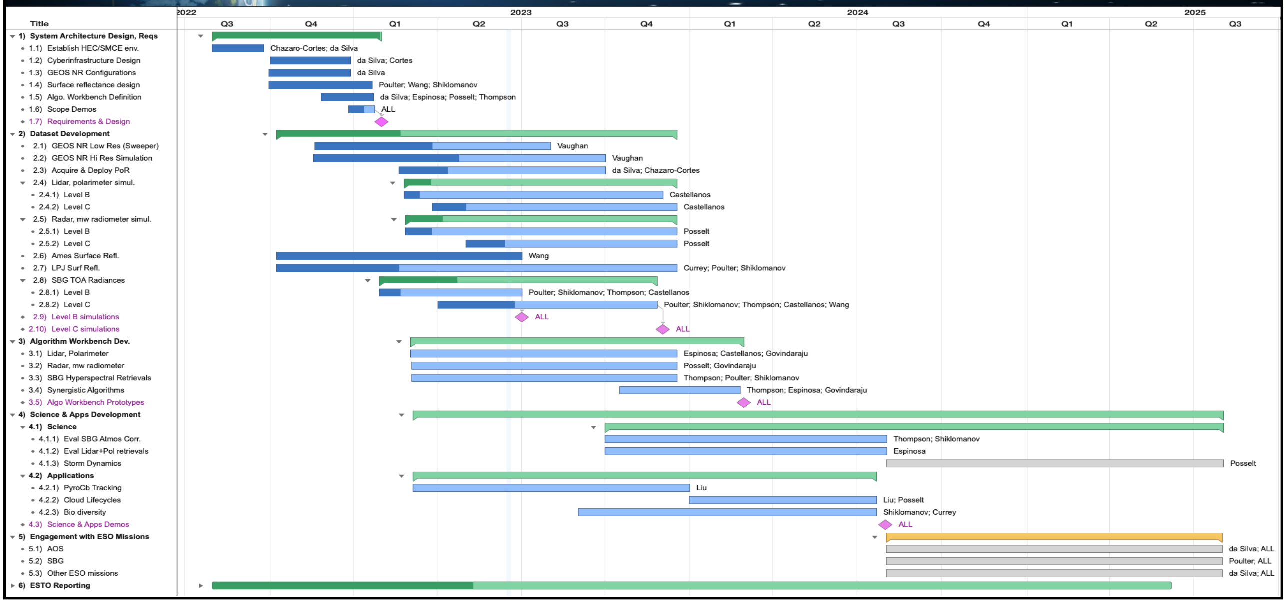
Extending these capabilities to simulated SBG/AOS data will accelerate algorithm development trade studies and provide the end user a sneak preview of that the ESO measurements can provide.

Plan Forward

Complete tasks as outlined in Gantt chart including…

- Complete dataset acquisition:
  - Geostationary Level 1B and Level 2 (aerosols)
  - Interface LPJ/PROSAIL to high resolution GEOS atmosphere (1.5 km
  - Generate reflectances and synthetic retrievals from GEOS Nature Runs and Ames/LPJ surface reflectances:
    - AOS Instruments: Lidar, polarimeter, Ku and W/Ka radars and microwave radiometers
    - SBG radiometer using both Ames surface reflectance and LPL/PROSAIL
- Start/continue work on algorithm workbench for AOS and SBG instruments
  - Exploring adoption of SlideRule Earth technologies for cloud optimized data access (via H5Coro) and on-demand generation of customizable algorithms.
- Complete work on tracking of Pyro Cb Plumes application using *tobac* cloud object technology.

Project Schedule



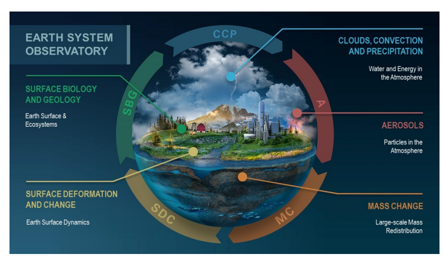
Final Remarks

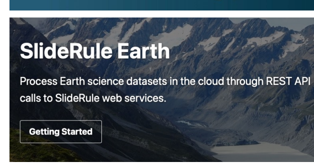
- The project is picking up momentum after some minor setbacks
- Challenges encountered during first year:
  - Generation of Storm Resolving Nature Run somewhat affected by delays in Clune's AIST project (Presentation Later Today)
  - Applications coordinator has accepted another position within GSFC; alternate identified – opportunity for injection of object identification technology (*tobac)*
  - HEC engineer provided by NCCS only available at 50%; others stepping in to fill the gap (see Team slide).
- Very good chemistry among team participants from different centers and disciplines
  - Great potential for AOS-SBG synergisms

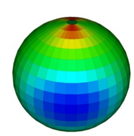
Presentation Contents

Infusions and Collaborations

- **Infusion:** AOS and SBG missions
  - Algorithm Workbench along with simulated data will be provided to the AOS and SBG missions
  - It helps that lead SBG/AOS scientists are co-investigators in this proposal
  - Full engagement with all ESO missions are planned for a Year 3 extension.
- **Collaborations:**
  - Leveraging Nature Run developments by Clune's AIST Project where PI da Silva is the Science PI.
  - Started conversation with SlideRule Earth group about adopting portions of their technology for cloud-optimized data access and algorithm workbench.
  - NASA centers involved: GSFC, JPL, ARC; leveraging MACIE airborne activity with strong LaRC participation (Gao Chen).
  - NCAR (ESMF core team): optimization of ESMF regridding (primarily via Clune's proposal)
  - Through connection with Clune's proposal, collaboration with the 2017 Decadal Survey Incubator (DSI) PBL including multiple investigators from GSFC, JPL, and LaRC. AOS lidar simulations of potential interest for DSI PBL. Co-I Posselt is a member of the DSI PBL team, being involved in several OSSE activities







Presentation Contents

- ✓Background and Objectives
- ✓Technical and Science Advancements
- ✓Summary of Accomplishments and Future Plans
- ✓Actual or Potential Infusions and Collaborations
- ○Publications - List of Acronyms

Publications

- Wang, W., Dungan, J., Genovese, V., Shinozuka, Y., Yang, Q., Liu, X., et al. (2023). Development of the Ames Global Hyperspectral Synthetic Dataset: Surface Bidirectional Reflectance Distribution Function. *Journal of Geophysical Research: Biogeosciences*, 128, e2022JG007363. https://doi.org/10.1029/2022JG007363

List of Acronyms

Acronym	Description
4SAIL	4-stream Scattering by Arbitrary Inclined Leaves RTM
AOS	Atmosphere Observing System
C576	Cubed-sphere with 5760 gridcells per edge. ~1.5 km global resolution
DSI	Decadal Survey Incubator
ESO	Earth System Observatory
ESMF	Earth System Modeling Framework (ESTO funded)
GEOS	Goddard Earth Observing System Earth System Model
ISOFIT	Imaging Spectrometer Optimal Fitting
LPJ	Lund-Potsdam-Jena (land surface model)
MC	Mass Change
NR	Nature Run
OSSE	Observing System Simulation Experiment
PROSAIL	Combination of PROSPECT and 4SAIL
PROSPECT	Propriétés Spectrales RTM
SBG	Surface Biology and Geology
SDC	Surface Deformation and Change

Glossary: OSSE Levels

- **Level A**: gridded properties
  - **Level A1**: gridded geophysical variables (GVs) from a *Nature Run (NR)*
    - In JEDI parlance these are called *GeoVals.*
  - **Level A2**: gridded surface characteristics, e.g., BRDF kernels, surface reflectances, or from high resolution land surface model
- **Level G**: geolocation file with coordinates, viewing and solar angles.
- **Level B**: Geophysical variables at *observation locations*
  - Input for radiative transfer calculations
  - **Level B1**: GVs from NR; sampling of Level A1 at observation locations
  - **Level B2**: input surface characteristics; sampling of Level A2 at obs locations
  - **Level B3**: geophysical variables used for validation of retrievals
- **Level C**: simulation of observables at observation location
  - Can include instrument model with random and systematic (calibration) errors.
  - **Level C1**: Single scattering properties
  - **Level C2**: Result of radiative transfer calculations, e.g., reflectances, DoLP
- **Level D**: retrieval simulations
- **Level E:** assimilation of synthetic measurements

Digital Twin Infrastructure Model for Agricultural Applications

Rajat Bindlish (PI, NASA GSFC)

Sujay Kumar, Pang-Wei Liu, Shahryar Ahmad, Jessica Erlingis Lamers (NASA GSFC); Alex Ruane, Meijian Yang, Luke Monhollon (NASA GISS); Zhengwei Yang (USDA NASS); Gary Feng, Yanbo Huang (USDA ARS)

AIST-21-0056 Annual Technical Review
June 23, 2023

Team listing: NASA GSFC, NASA GISS, USDA NASS, USDA ARS

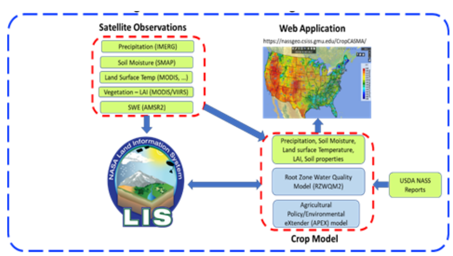
1

Digital Twin Infrastructure Model for Agricultural Applications

PI: Rajat Bindlish, NASA Goddard Space Flight Center

Objective

We propose the development of a Digital Twin Agriculture prototype by integrating land/hydrology process models, agricultural models, and remote sensing information. The development of the agricultural digital twin infrastructure will allow us to assess the socio-economic impact from naturally occurring and human activities on agriculture and food security.



Digital Twin Architecture for Agriculture

Approach

- Establish an Agricultural Digital Twin framework.
- Assimilate high-resolution remote sensing observations in the NASA's Land Information System (LIS).
- Develop a crop model operational framework.
- Couple NASA's LIS. With the crop model.
- Conduct 'what if' scenario investigations.
- Develop web application to disseminate results of the study.

Key Milestones

• Requirements, architecture, design	12/22
• Crop model development	12/23
• Couple crop model with NASA's LIS and Agriculture DT	06/24
• Web application development	03/25
• Scenario investigations	06/25

Co-Is/Partners: NASA GSFC, USDA ARS, USDA NASS

TRL_{in} = 3 $TRL_{current}$ = 6

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- List of Acronyms

Background / Objectives

- We aim to develop a Digital Twin Agriculture prototype by integrating land/hydrology process models, agricultural models, and remote sensing information. The development of the agricultural digital twin infrastructure will allow us to assess the socio-economic impact from naturally occurring and human activities on agriculture and food security.
- The current crop models do no adequately represent land surface hydrology
- LIS current does not include crop modeling but has state-of-the-art land surface models
- “Agricultural Twin” will provide a state-of-the-art agriculture model that will provide capability for food security monitoring

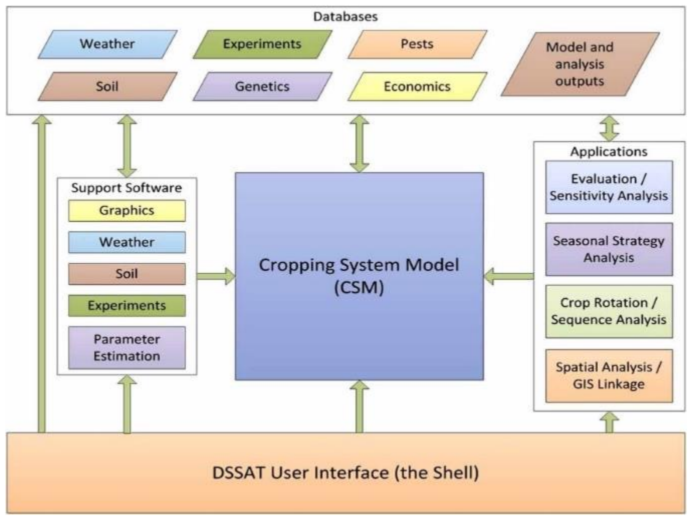
Specific Objectives

- Establish a digital twin framework that enables the NASA remote sensing data products and land surface model products to be directly coupled with or assimilated into the crop growth model
- Assimilate high-resolution remote sensing inputs (precipitation, temperature, soil moisture, snow water equivalent, ground water, leaf area index) through the NASA Land Information System (LIS) to estimate land surface variables (water and energy fluxes) at daily time scales
- Implement crop growth models [Agricultural Policy/Environmental eXtender (APEX), Root Zone Water Quality Model (RZWQM2) and Decision Support System for Agrotechnology Transfer (DSSAT)] to estimate crop growth states, biomass, and crop yield under long-term weather conditions and projected future climate scenarios
- Implement Bayesian Neural Network (BNN) model to predict final county level crop yield by using NASS historical yield reports, yield, biomass, and phenology outputs from APEX/ DSSAT model, and other variables from LIS
- Develop tools to conduct ‘what if’ investigations to provide agricultural guidance
- Develop capability for disseminating non-confidential crop progress data, biomass and crop yield maps using an operational web application

Presentation Contents

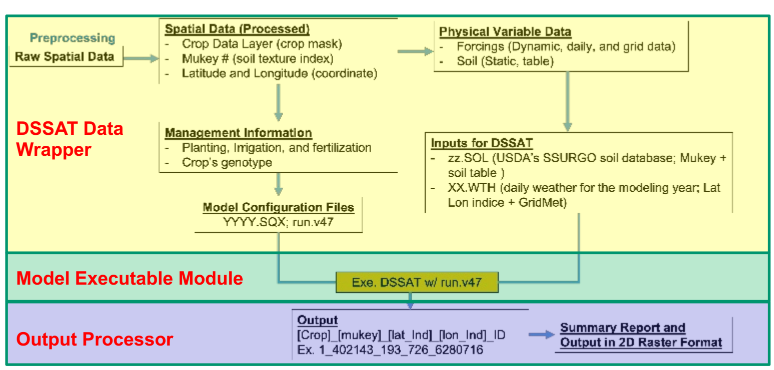
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- List of Acronyms

DSSAT Crop Model



Crop Model Development 1/3

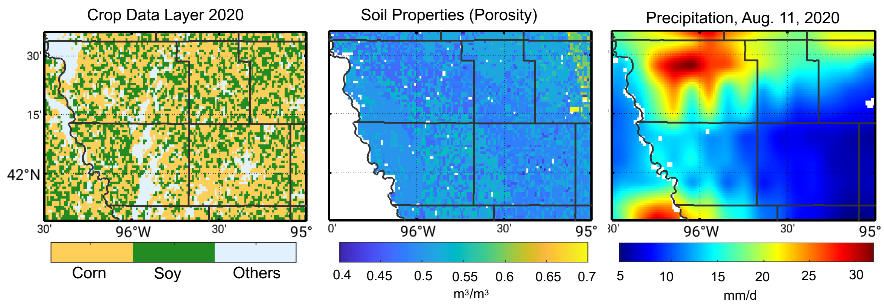
- A crop growth model, Decision Support System for Agrotechnology Transfer (DSSAT), was selected and developed for the project.
  - A data wrapper was developed to ingest input variables including crop type, soil properties, and weather forcings for the domain of interest.
  - Executable and output processor module are developed to run model in a large domain.



DSSAT Model Flowchart

Crop Model Development 2/3

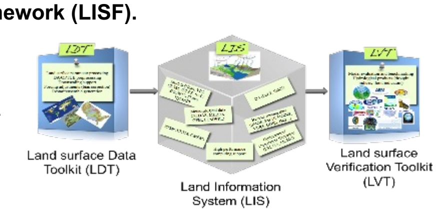
- DSSAT was run in four agricultural dominant counties, Woodbury, Ida, Monona, and Crawford in Iowa for model test.
- **Crop data layer:**
  - Crop data layer from USDA NASS are used as crop mask.
  - Corn and soybean are targeted crops.
- **Soil properties:**
  - USDA's Soil Survey Geographic Database (SSURGO) is used.
- **Weather forcings:**
  - Precipitation from GPM IMERG, and solar radiation, temperature (max and min), wind speed from MERRA-2 are used to drive the crop model.



Configuration of NASA's LIS for Coupling 1/3

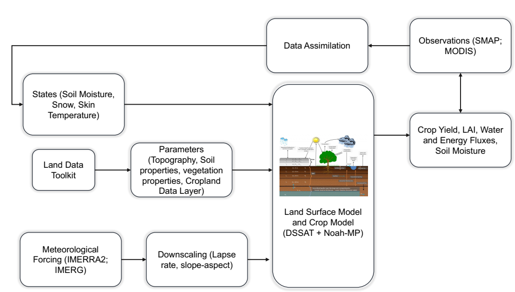
NASA GSFC's Land Information System Framework (LISF).

- Land surface Data Toolkit (LDT):
  - Tool to preprocess input variables
- Land Information System (LIS):
  - Implementation of land surface model and DA.
- Land surface Verification Toolkit (LVT):
  - Tool to evaluate model estimates.



Model Configuration for Coupling W/ DSSAT

- Land surface model: Noah-MP 4.0.1
- Grid Spacing: 0.01 x 0.01 degree
- Soil Layers: 0 - 10 cm, 10 - 40 cm, 40 - 100 cm, 100 - 200 cm
- Land cover: MODIS
- Soil texture: 1 km STATSGO-FAO
- Elevation: SRTM
- Greenness fraction: MODIS/NOAA-NCEP
- Albedo: MODIS/NOAA-NCEP
- Forcing data: MERRA2 and IMERG

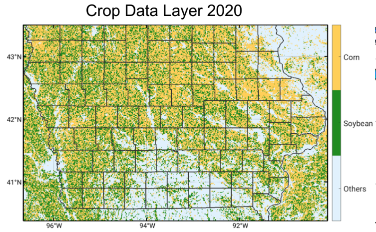


Configuration of NASA's LIS for Coupling 2/3

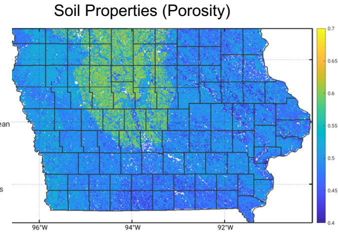
LDT development for crop data layer and gSSURGO soil datasets

- Domain definition module
- Scalable module
  - Nearest neighbor
  - Mode
- Mukey value of soil database is related to NASA's SSURGO Table
- CDL and soil properties were aggregated to 1 km for the experiment.

Crop Data Layer 2020



Soil Properties (Porosity)



USDA's SSURGO

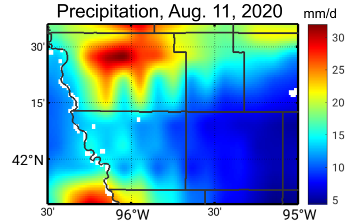
	AREASYMBOL	MUKEY	SLCL_5	SLCL_150	SLCL_180	SLCL_30
0	IA191	1000027	21	26	26	21.5
1	IA191	1008290	9	4	4	11
2	IA191	1008292	6	4	4	6
3	IA191	1008301	21	-99	-99	21
4	IA191	1008304	9	4	4	11
5	IA191	1008306	9	4	4	11
6	IA191	1008421	22	-99	-99	19
7	IA191	1008428	-99	20.299999237060547	20	0
8	IA191	1008429	28	28	28	28
9	IA191	1008430	18.5	24	24	19.5
10	IA191	1014251	22	24	21.399999618530273	22
11	IA191	1014259	18.5	56	56	21.100000381469727
12	IA095	1144372	24	32	32	26
13	IA095	1144373	22.5	22.100000381469727	22	22.5
14	IA095	1144375	23	23	22.5	21.299999237060547
15	IA095	1144376	22	22	22	24
16	IA095	1144377	22	22	22	24
17	IA095	1144379	5	4	4	3
18	IA095	1144380	20	26	26	24.100000381469727

Configuration of NASA's LIS for Coupling 3/3

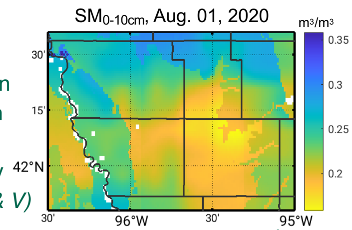
Input Parameters:
vegetation class
vegetation greenness/LAI
soil type
elevation

Required Forcing Fields:
total precipitation
convective precipitation
downward shortwave radiation
downward longwave radiation
near surface air temperature
near surface specific humidity
near surface wind speed *(U & V)*
surface pressure

Precipitation, Aug. 11, 2020



$SM_{0\text{-}10cm}$, Aug. 01, 2020



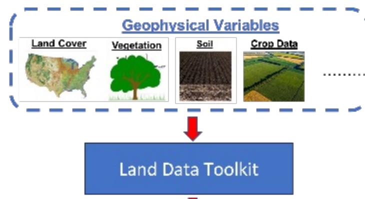
Assimilated Data Fields:
snow cover
snow depth or water equivalent
soil moisture
water storage anomaly
leaf area index
irrigation intensity
surface temperature

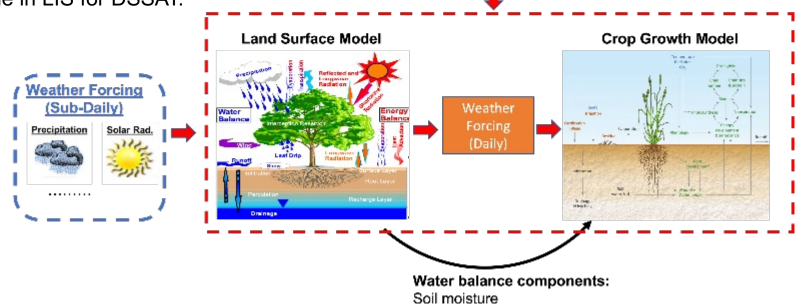
Summary of Output Fields:
soil moisture in each layer
groundwater storage
soil temperature in each layer
surface and subsurface runoff
evaporation
transpiration
latent, sensible, and ground heat fluxes
rainfall
net shortwave and longwave radiation

Couple Crop Model With NASA's LIS

Agriculture DT

- Land-Agriculture Information System (LAIS) is built based on coupling LIS and DSSAT.
- The phase I development was built upon loose coupling soil moisture between the two systems and has been implemented.
- Weather forcings are aggregated to daily scale in LIS for DSSAT.





Complexity of Coupling LIS W/ DSSAT

Model mechanism inconsistency

- LSM in LIS is designed to run spatially in a 2D domain for each time step; while DSSAT model is running per pixel over a period of time.
  - Data wrapper was developed to circle around such an inconsistency for loose coupling the two systems.
  - A restart module for DSSAT is under development for tight coupling.

Model physics inconsistency

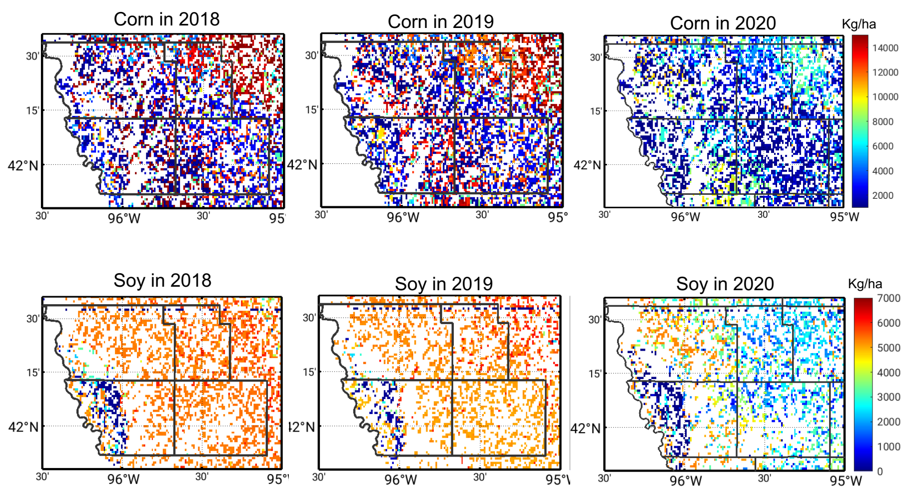
- Soil layers are treated differently in LSM and DSSAT (homogeneous compared to heterogeneous profiles).
  - Soil profiles are layered consistently through preprocessing to four layers, consistent to LSM.

- Inconsistent soil properties input in LSM and DSSAT (soil texture compared to soil property variables).
  - A soil profile module is being developed to unify soil input variables in DSSAT for LIS LSM.

- Weather forcing scale inconsistency in LSM and DSSAT (sub-daily compared to daily).
  - Weather forcing is aggregated in LIS to daily scale for loose coupling w/ DSSAT.

Couple Crop Model With NASA's LIS

- Crop yields for corn and soy in 2018-2020 are estimated using LIS-DSSAT model.



Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- List of Acronyms

Summary of Accomplishments and Future Plans

Current status of project
- Development of Crop Model
  - Updated DSSAT model to Version 4.8.
  - Developed and implemented a new module in DSSAT to integrate external soil moisture for model coupling.
  - Developed and modified data wrapper to consider crop rotation and multiple year of model spin-up.

- Couple crop model with NASA's LIS and Agriculture DT
  - Developed data reader and writer for CDL and gSSURGO soil database in LDT.
  - Developed a framework to loosely couple NASA's LIS and DSSAT model.
  - Implemented crop simulation for corn and soybean in selected counties in Iowa.

- Web application development
  - Tested USDA's Crop Condition and Soil Moisture Analytics (Crop-CASMA) to ingest simulation output variables, such as crop yield, phenology, vegetation, and water balance parameters.

- Scenario investigations
  - Understood the simulation inconsistency between DSSAT alone and LIS-DSSAT models.

Summary

- The project is progressing well
- The integration of the two modeling systems: Land surface hydrology (LIS) and Crop Model (DSSAT) is in progress
- The value of precip and land surface hydrology variables (soil moisture, soil temperature) was demonstrated
- Enable better crop yield forecasts that use antecedent surface observations
- This project will enable the first integration of assimilated remote sensing observations in the land surface hydrology with a crop model

Summary of Accomplishments and Future Plans

Future plans

- Development of Crop Model
  - Develop module for coupling vegetation component, eg. LAI, for model coupling.
  - Develop data assimilation schemes using remotely sensed data for DSSAT model.
- Couple crop model with NASA's LIS and Agriculture DT
  - Investigate the impact of model inconsistency in soil profiles.
  - Develop plug-in module to integrate DSSAT model into LISF.
  - Develop a tight coupling scheme for LIS-DSSAT framework.
- Web application development
  - Develop a data module to ingest direct outputs from LIS-DSSAT model.
- Scenario investigations
  - Investigate impacts of coupling models on climatology on soil moisture, temperature, LAI, and ET.
  - Implement DA using remotely sensed datasets in the coupling framework and assess its impacts on the agriculture system.
  - Investigate 2016 and 2017 drought in Northern Great Plaint and 2019 flood in mid-west.

ESTO Earth Science Technology Office

19

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- List of Acronyms

ESTO Earth Science Technology Office

20

Actual or Potential Infusions and Collaborations

- Alex Ruane (NASA GISS) has agreed to be a member of the project team. Alex has extensive experience in crop modeling and leads the AgMip project.
- The team includes USDA NASS investigators who are very interested in infusing this technology
- Met with NASA water resources and agriculture applied science program
- Working with NASA Acres project at UMD – NASA domestic agriculture consortium

ESTO Earth Science Technology Office

21

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- List of Acronyms

ESTO Earth Science Technology Office

22

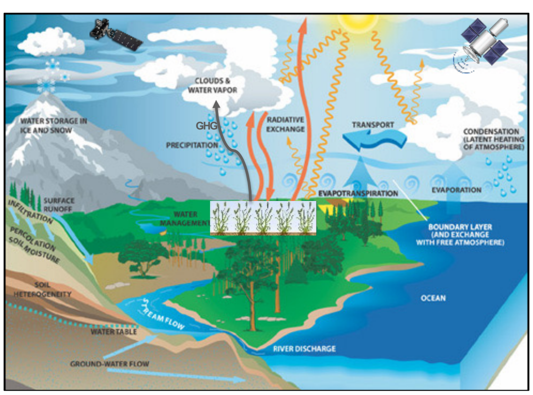
List of Acronyms

- ARS — Agricultural Research Service
- CDL — Crop Data Layer
- DA — Data Assimilation
- DSSAT — Decision Support System for Agrotechnology Transfer
- DT — Digital Twin
- GISS — Goddard Institute for Space Studies
- GSFC — Goddard Space Flight Center
- GSSURGO — Gridded Soil Survey Geographic
- LDT — Land Data Toolkit
- LIS — Land Information System
- NASS — National Agricultural Statistics Service

ESTO Earth Science Technology Office

23

LIS-Agriculture DT on Ecosystem



Conceptual Diagram of LIS and Agriculture DT

Capabilities of LIS-Agriculture DT

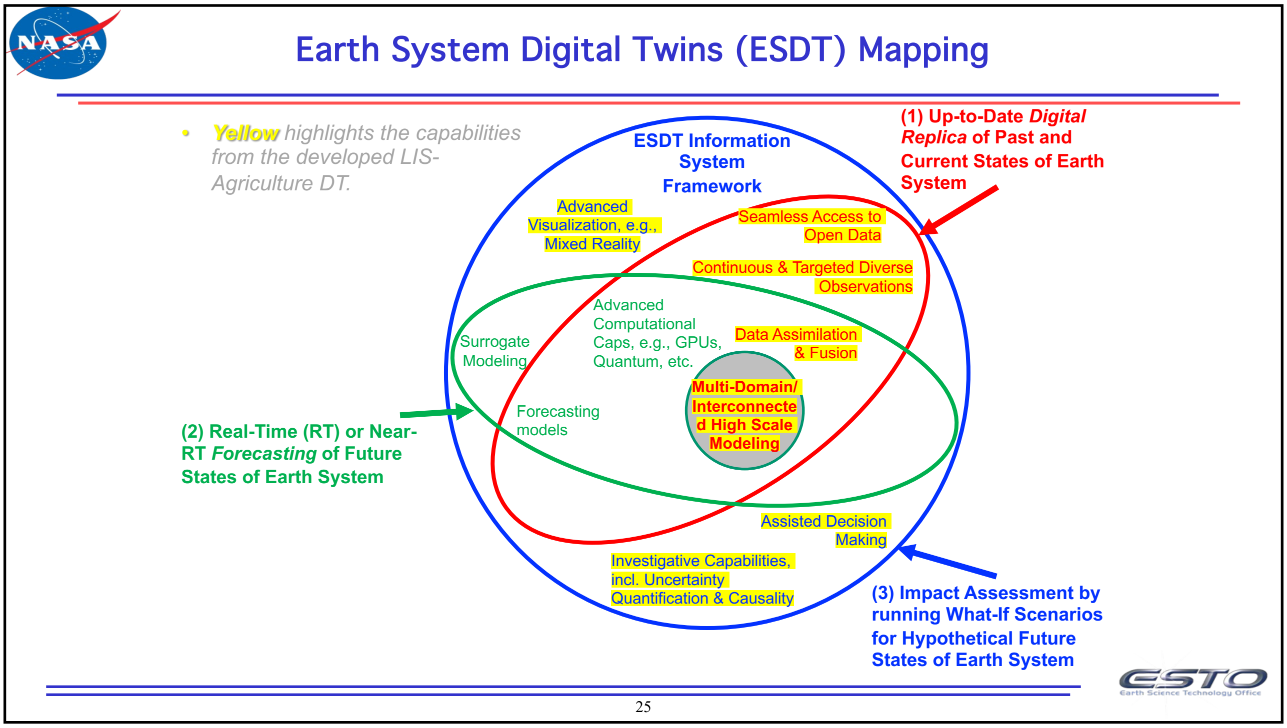
- Couple land surface processes with agricultural crop growth processes on croplands.
- Integrate NASA's satellite observations to improve estimates from the model such as hydrological variables and crop yields in croplands.

Contributions of LIS-Agriculture DT to Ecosystem

- Understand impacts of agricultural systems on water cycle to enhance water management capability.
- Investigate freshwater pollution due to fertilization in the agricultural system.
- Investigate green house gas emission (GHG) due to agricultural activities and its impacts on climate change.

ESTO Earth Science Technology Office

24


Earth System Digital Twins (ESDT) Mapping
Yellow highlights the capabilities from the developed LIS-Agriculture DT.
ESDT Information System Framework
(1) Up-to-Date Digital Replica of Past and Current States of Earth System
Advanced Visualization, e.g., Mixed Reality
Seamless Access to Open Data
Continuous & Targeted Diverse Observations
Advanced Computational Caps, e.g., GPUs, Quantum, etc.
Surrogate Modeling
Data Assimilation & Fusion
Multi-Domain/ Interconnected High Scale Modeling
Forecasting models
(2) Real-Time (RT) or Near-RT Forecasting of Future States of Earth System
Assisted Decision Making
Investigative Capabilities, incl. Uncertainty Quantification & Causality
(3) Impact Assessment by running What-If Scenarios for Hypothetical Future States of Earth System
ESTO
Earth Science Technology Office
25

Development of a next-generation ensemble prediction system for atmospheric composition

Christoph Keller (PI, MSU/NASA GMAO)
Jennifer Sleeman (Co-PI, JHU/APL)
Patricia Castellanos (Co-I, NASA GMAO)

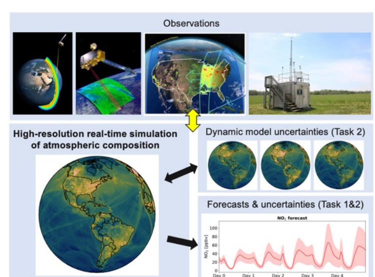
AIST-21-0024 annual review
2023-06-23

Development of a next-generation ensemble prediction system for atmospheric composition

PI: Christoph Keller, MSU / NASA GMAO; Co-PI: Jennifer Sleeman, JHU / APL

Objective

- Develop a novel and practical framework to conduct high-resolution, real-time simulations of reactive gases by (a) reducing compute requirements for atmospheric chemistry simulations and (b) developing machine learning models to efficiently predict model uncertainties.
- This Framework will be applied to the NASA GEOS Composition Forecast system (GEOS-CF) first and transferred to NOAA's air quality forecasting system in year 3.
- Project goals include 3-fold speedup of short-term simulations (<14 days); doubling of horizontal model resolution; increase in forecast length from 5 to 10 days.



Approach

Enhance current GEOS model capabilities by:

Task 1: Enable accelerated simulations of reactive trace gases in GEOS by implementing machine learning emulator for atmospheric chemistry and simplify model parameterization of tracer transport.

Task 2: Develop deep learning (DL) model to dynamically predict model uncertainties of key trace gases, such as nitrogen dioxide (NO_2), ozone (O_3), and aerosols.

Co-PI: Jennifer Sleeman (JHU)
Co-I: Patricia Castellanos (NASA GMAO)

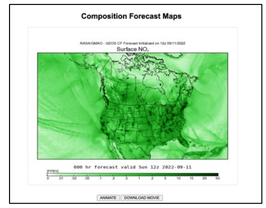
Key Milestones

- 02/23: Chemistry emulator implemented (TRL 5)
- 06/23: Simplified transport implemented (TRL 5)
- 08/23: DL prototype (TRL 4)
- 02/24: New GEOS-CF workflow complete (TRL 6)
- 08/24: Technology transfer to NOAA (TRL 7)

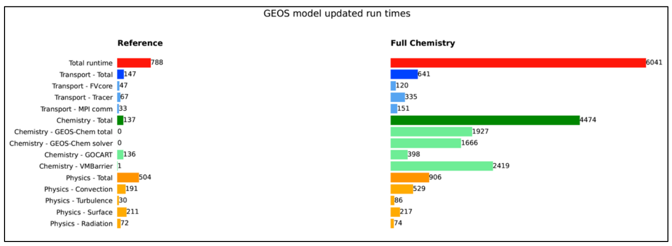
Current TRL = 4

Simulating atmospheric composition is expensive

The NASA **GEOS Composition Forecast System** (GEOS-CF) produces daily global forecasts of atmospheric composition. However, GEOS-CF is computationally very expensive, which restricts the range of applications.



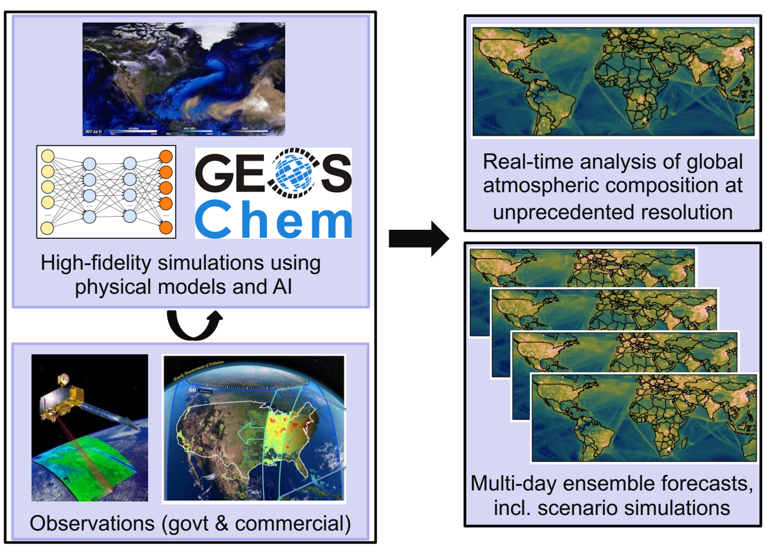
https://fluid.nccs.nasa.gov/cf/classic_geos_cf/



- ~0.25 degrees horizontal resolution (global), 72 vertical levels
- >200 chemical species, >700 reactions
- 3’580 CPU’s

ESTO Earth Science Technology Office

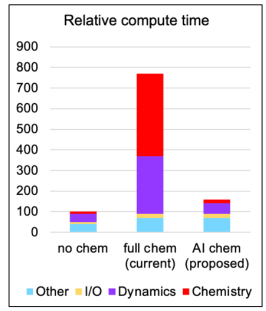
Project objective: accelerate real-time simulation of atmospheric composition



ESTO Earth Science Technology Office

Project overview

- **Tasks:**
  - Task 1: **Accelerated version of GEOS-CF** suitable for short-term applications
  - Task 2: Devise an **ensemble prediction framework** using deep learning (DL)
  - Task 3: **Apply the framework** to GEOS-CF and NOAA's system (optional 3rd year)



- **Science / Technology advancements:**
  - Deliver DL framework for GEOS model system
  - Speed up (short-term) model simulations of atmospheric composition by factor 3-5
  - Increase GEOS-CF model resolution by 2x
  - Double GEOS-CF forecast length to 10 days

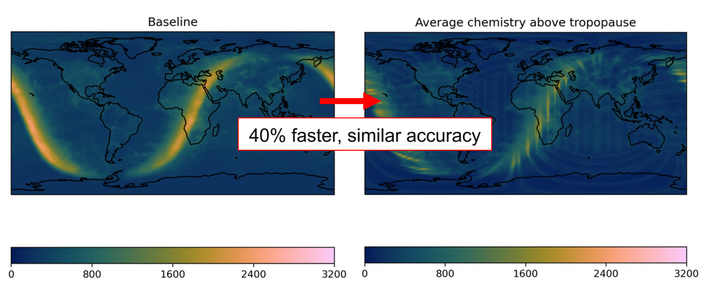
- **Relation to NASA R&A and Applications:**
  - Real-time simulation of atmospheric composition (Atmospheric Composition; Carbon Cycle; Health & Air Quality)
  - Improved air quality forecasts (Health & Air Quality; Fires)

ESTO Earth Science Technology Office

5

Task 1: Accelerate chemical solver

- **Goal:** faster version of GEOS-CF, suitable for forecasts and scenario simulations
- **Activity:** reduce terminator chemistry runtime:
  - Grid-cell averaging
  - Flexible adjustment of error tolerances
- **Motivation:** twilight regions account for ~10% of global domain, but consume 33% of computation
- **Team:** Christoph Keller, Obin Sturm, Mike Long, Patricia Castellanos

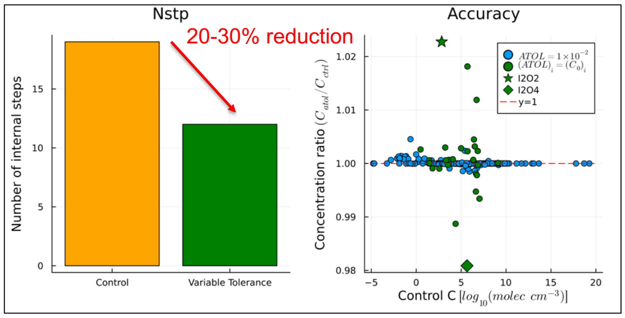


ESTO Earth Science Technology Office

6

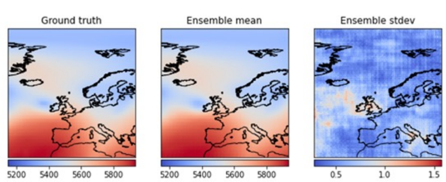
Task 1: Dynamical error adjustment

- Identified novel way of speeding up the numerical calculation of chemistry in the KPP software package (by optimizing error tolerances)
- Efficiency tests and implementation into GEOS-CF currently underway
- Initiated collaboration with KPP developers

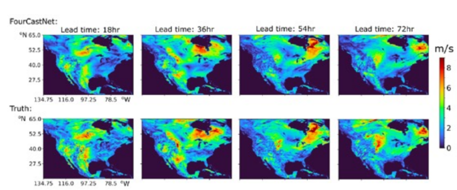


Overview Task 2: Deep Learning Models

- **Goal:** develop deep learning model suitable for GEOS-CF ensemble predictions
- **Activity:** test two DL approaches tested for weather forecasting:
  - c-GAN (Bihlo, 2021)
  - Fourier-based neural network (FourCastNet; Pathak et al., 2022)
- **Motivation:** Ensemble simulations using the numerical model is computationally too expensive, deep learning models can potentially solve this problem
- **JHU APL Team:** Jennifer Sleeman, Chris Ribaudo, David Chung, Mimi Szeto, Marisa Hughes



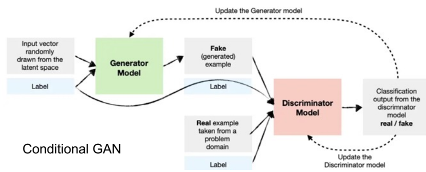
Bihlo, 2021

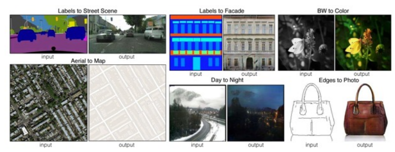


Pathak et al., 2021

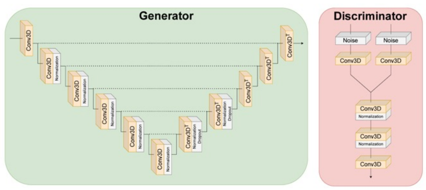
Task 2: Conditional Generative Adversarial Networks (cGANs) for Model Ensembling

- Bihlo (2021) used the Monte Carlo Dropout methodology and cGANs for weather forecasting and ensembling
- The cGAN method is an ensemble of 100 realizations of the deep learning model
- It does not vary initial conditions (relies on drop-out to vary model after model is trained)

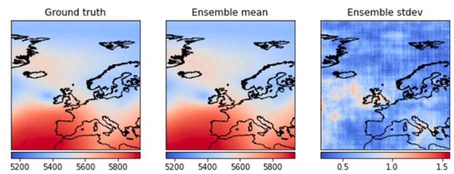




Based on Image-to-Image Translation (Isola 2017)



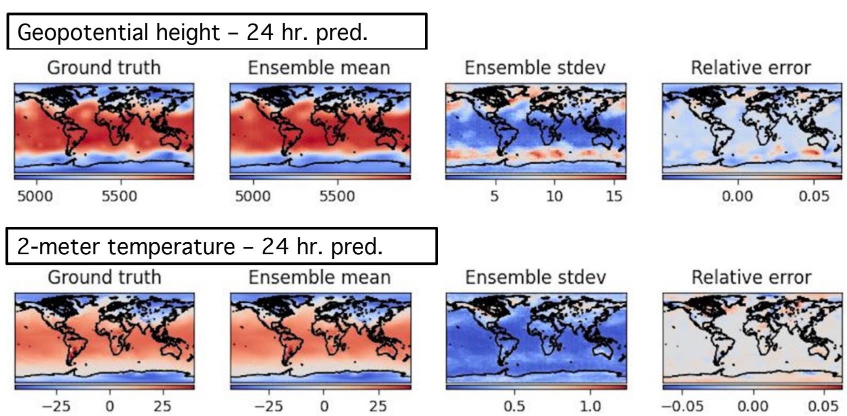
Brecht, Bihlo (2023) Vid-to-vid Architecture U-Net Generator (left) and Discriminator (right)



Bihlo's cGAN method showed to produce realistic weather model ensembles with a mean that accurately reflects the numerical model.

Bihlo 2021: https://doi.org/10.1016/j.neunet.2021.02.003
Brecht & Bihlo 2023: https://doi.org/10.1029/2022GL101452

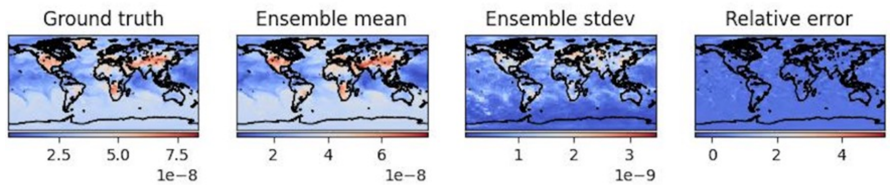
Able to reproduce published cGAN results using GEOS output



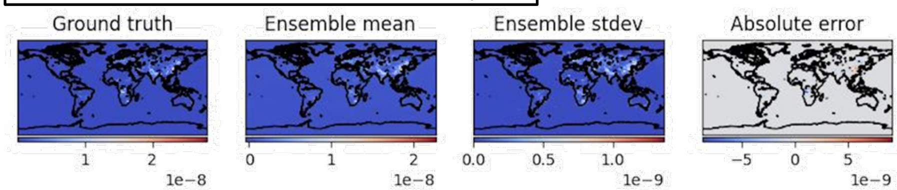
- 500 hPa geopotential height, two-meter temperature
- Training set: 11 months
- Test set: 1 month (Aug 2019)

Predicting pollutant concentrations using cGAN: promising initial results

O_3 concentration (summer 2019) – 24 hr. pred.

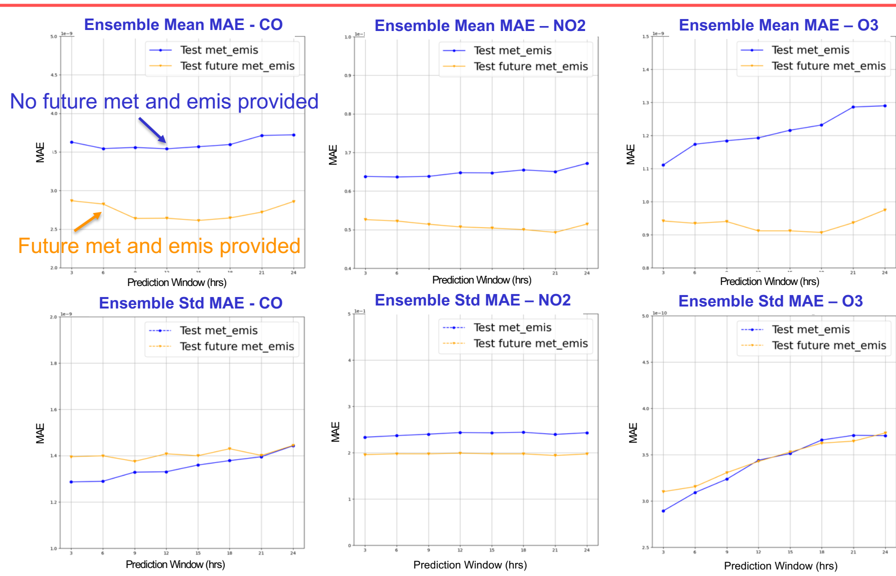


NO_2 concentration (summer 2019) – 24 hr. pred.



ESTO Earth Science Technology Office

Improved forecasts when providing projected meteorology and emissions



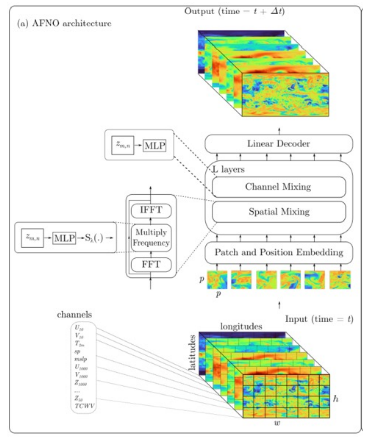
ESTO Earth Science Technology Office

FourCastNet for Model Emulation and Uncertainty Estimation Research

- FourCastNet is a deep neural network that learns to forecast 2D atmospheric variables globally

Vertical Level	Variables
Surface	U_{10}, V_{10}, T_{2m}, sp, $mslp$
$1000hPa$	U, V, Z
$850hPa$	T, U, V, Z, RH
$500hPa$	T, U, V, Z, RH
$50hPa$	Z
Integrated	$TCWV$

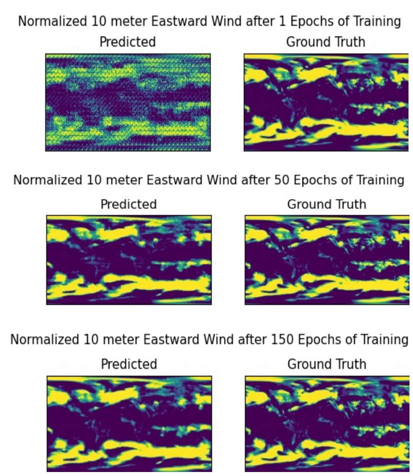
- Uses Fourier Neural Operator (FNO) combined with vision transformer backbone
  - FNO shown to perform well in modelling complex PDE systems
- Pathak et al. (https://arxiv.org/abs/2202.11214)



Training a new FourCastNet model using NASA GEOS output

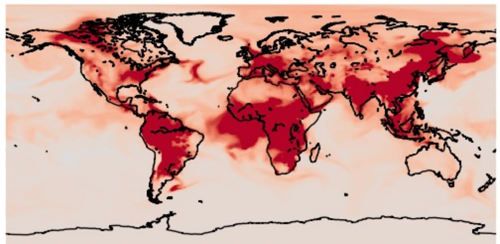
- Traditionally trains on ERA reanalysis output
  - 0.25 degrees resolution
- Reconfigured FourCastNet code to train on NASA GEOS output
  - 1 degrees resolution (initially)
  - 315 days of training data (Feb 21, 2020 – Dec 31, 2020)
  - Output from training

Able to train our own FourCastNet on NASA GEOS output
Verified framework and FourCastNet data requirements for GEOS

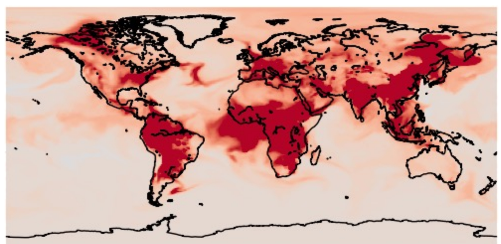


FourcastNet trained on composition data

CO Ground Truth

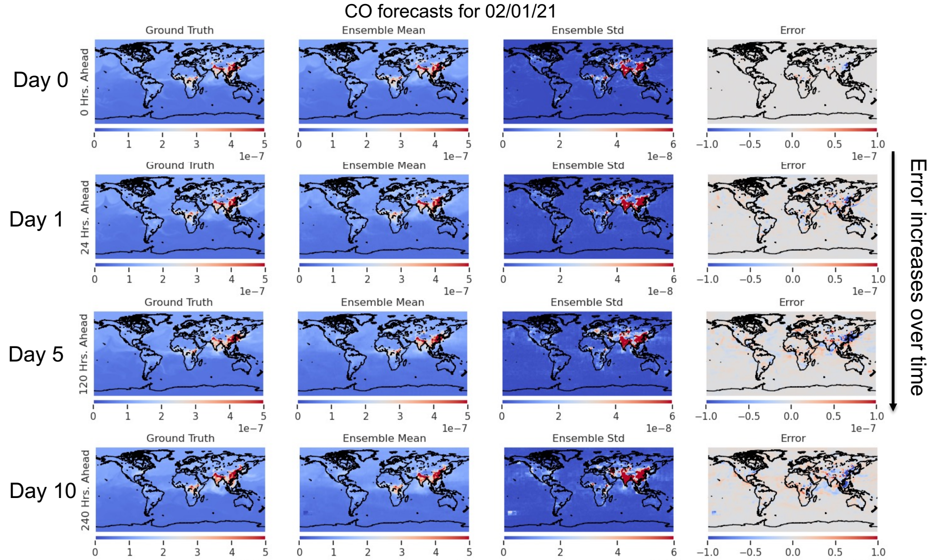


CO Prediction



FourcastNet ensemble learning: 10 day predictions

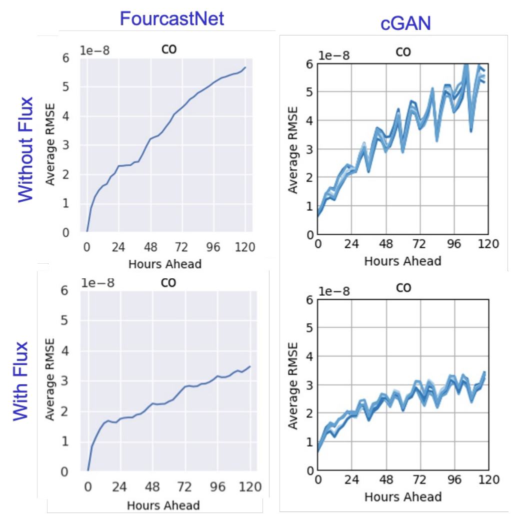
CO forecasts for 02/01/21



Sensitivity to emissions (FourCastNet and cGAN)

- Generated comprehensive suite of ensemble training dataset (dataset generation ongoing)
- Updated both AI models (cGAN and FourCastNet) to provide more flexibility in choice of model architecture and input features.
- Like-for-like comparison of cGAN vs. FourCastNet to better understand advantages and limitations of both approaches (see figure).

Early experiments show a smaller average RMSE when emissions are included with concentrations as input for both FourcastNet and cGAN.



ESTO Earth Science Technology Office

17

Current State

- **Task 1 (Accelerated version of GEOS-CF):**
  - Optimizing KPP compute costs via error tolerances
  - Testing this feature in GEOS-CF
- **Task 2 (DL model for ensemble predictions):**
  - Developed DL model prototypes based on c90 (1x1 degree) output:
    - cGAN model
    - FourCastNet model
  - Generated comprehensive set model ensembles

ESTO Earth Science Technology Office

18

Summary and Future Plans

Summary
- On schedule to deliver:
  - Accelerated version of GEOS-CF suitable for short-term simulations
  - Ensemble prediction capability using DL models
- First results using cGAN / FourCastNet are very encouraging
- KPP optimization via error tolerances is a new development, with potential benefit to the wider KPP community

Future Plans
- Task 1:
  - Finalize KPP error tolerances optimization
  - Implement and benchmark chemistry emulator in GEOS-CF
- Task 2:
  - Test modified DL models for chemistry application
  - Test ability to predict ensembles spread based on input perturbations
  - Fine-tune DL model and interface with GEOS-CF

Contact: christoph.a.keller@nasa.gov

Infusions and Collaborations

- Potential infusions
  - Bi-weekly meetings with NOAA model developers
- Ongoing collaborations
  - GEOS-Chem Support Team, Max-Planck Institute (KPP integration)
  - AIST-18-0011 (Randall Martin)
  - University of Southern California (Terminator Chemistry)
  - Atmospheric Chemistry and Dynamics Laboratory (GSFC Code 614)
  - NVIDIA & Lawrence Berkeley National Laboratory (FourCastNet)

Publications

Acronyms

List of Acronyms

- ACC Anomaly Correlation Coefficient
- AFNO Adaptive Fourier Neural Operator
- AQF Air Quality Forecast
- cGAN conditional generative adversarial network
- DL Deep Learning
- GEOS Goddard Earth Observing System
- GEOS-CF GEOS Composition Forecast
- GMAO Global Modeling and Assimilation Office
- IOA Index of Agreement
- JEDI Joint Effort for Data Assimilation Integration
- JHU/APL Johns Hopkins University / Applied Physics Laboratory
- KPP Kinetic Pre-Processor
- MSU Morgan State University
- NMB Normalized Mean Bias
- RMSE Root Mean Square Error
- RQE Relative Quantile Error
- TRL Technology Readiness Level

Development of the High Performance Implementation of GEOS-Chem (GCHP) to enable broad community access to high-resolution atmospheric chemistry modeling in support of NASA Earth Science

Randall Martin (PI, Washington University)
Daniel Jacob (Co-I, Harvard University)
Tom Clune (Co-I, NASA GSFC)
Christoph Keller (Co-I, NASA GSFC)
Sebastian Eastham (Co-I, MIT)

AIST-18-0011 Annual Technical Review
June 23, 2023

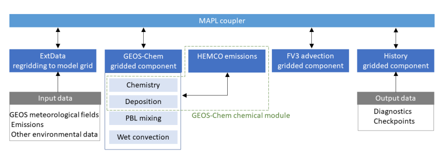
Bob Yantosca (Harvard), Jourdan He (WashU), Lizzie Lundgren (Harvard), Saptarshi Sinha (WashU), Dandan Zhang (WashU)

Augmentation to Development of the High Performance Implementation of GEOS-Chem (GCHP) to Enable Broad Community Access to High-resolution Atmospheric Chemistry Modeling in Support of NASA Earth Science

PI: Randall Martin, Washington University

Objective

- Further develop GCHP to support broad community access to high-resolution atmospheric chemistry modeling and chemical data assimilation, and including by recently funded AIST projects (i.e., Huang and Keller).
- Enhance community access to the cloud-enabled GCHP, a high-performance chemical transport model,
- Integrate the following technologies: high performance atmospheric chemistry model (GCHP); cloud computing; cloud pipelines; adaptive chemical solvers; smart load balancing; Modeling Analysis and Prediction Layer (MAPL);



Schematic of GCHP architecture including the GEOS-Chem chemical module used either offline as a CTM or online in an ESM, with interfaces managed through the MAPL coupler.

Approach

- Improve GCHP accessibility including on the cloud through demonstration and documentation of capabilities, a cataloging system to manage input data, and automated pipelines on the cloud
- Develop algorithms to increase GCHP computational performance through adaptive chemical solvers, aerosol microphysics, and smart load balancing
- Modularize GCHP chemistry and thermodynamics solver to enable interoperability

Key Milestones

• Adaptive chemical solvers	02/23
• Cataloging system to manage input data	05/23
• Automated pipelines on cloud	02/24
• Modularized GCHP components	02/24

Co-Is/Partners: Daniel Jacob, Harvard; Tom Clune, Christoph Keller, GMAO; Steven Barrett, Sebastian Eastham, MIT

TRL_{in} = 4 **$TRL_{current}$ = 5**

Presentation Contents

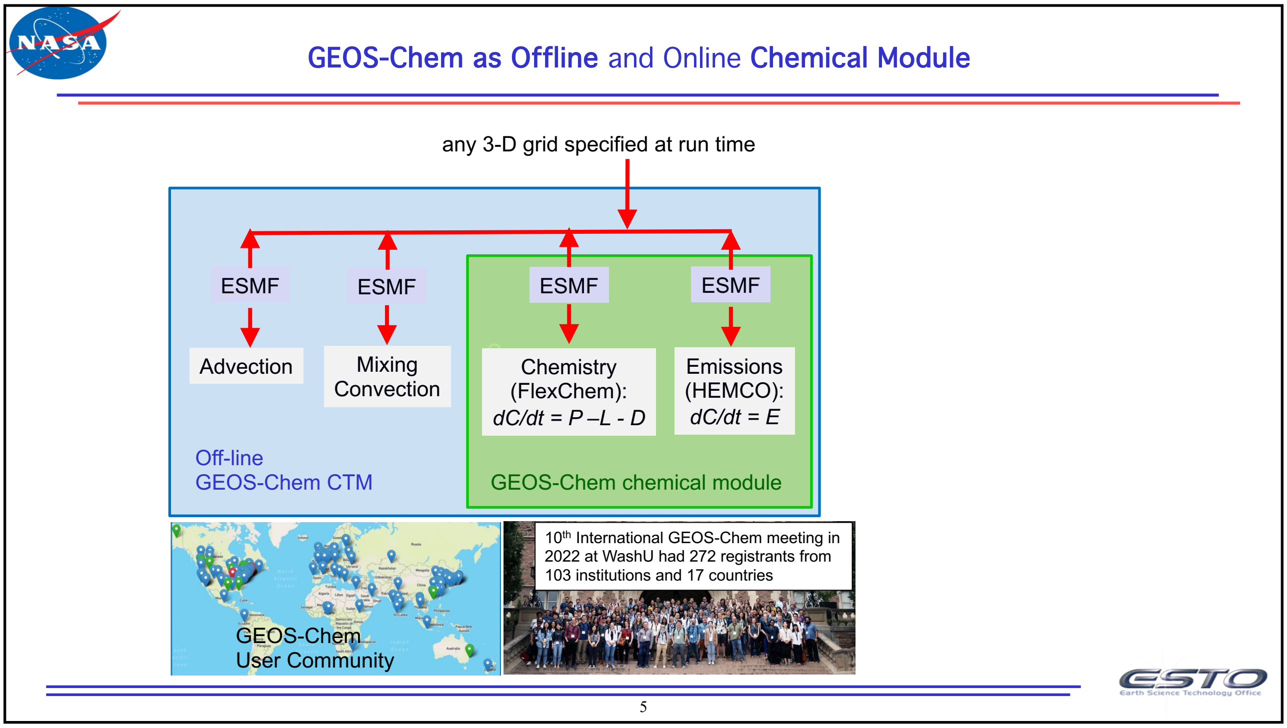
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

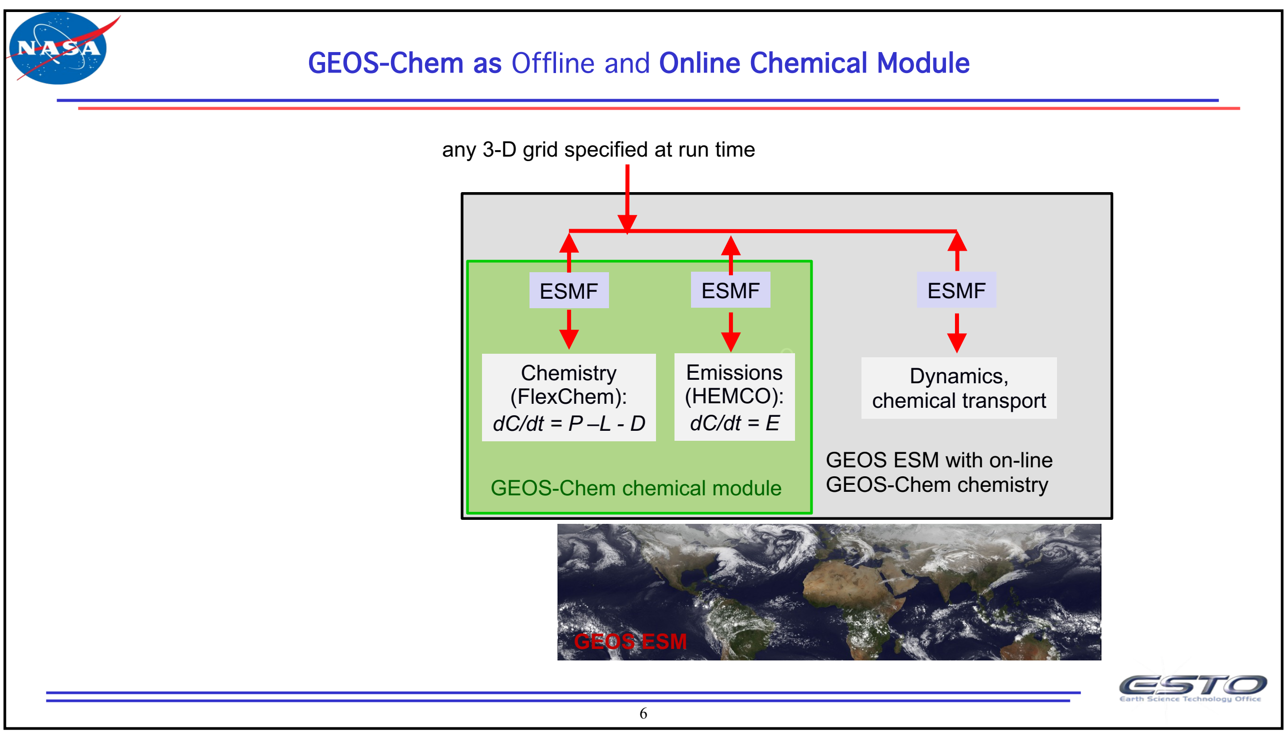
Background and Objectives

- Atmospheric chemistry models are needed to interpret observations and enable predictions for a range of scientific investigations including air quality, the carbon cycle, climate processes, and extreme events
- These models must be able to operate not only online as components of Earth system models (ESMs) but also offline, using meteorological data as input, because the atmospheric chemistry community relies on the more easily accessible offline version for model development and applications.

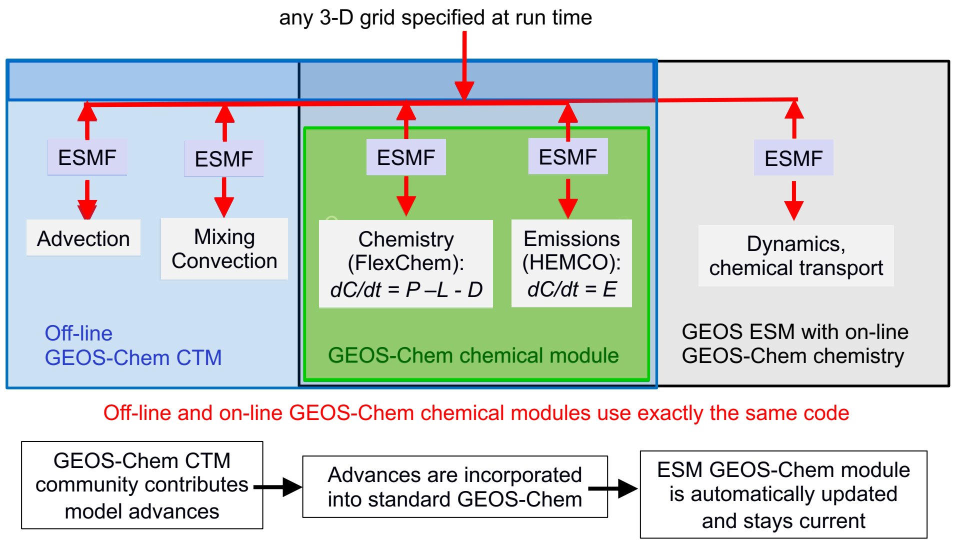
- This project helps meet the Research and Applications science goals for several cross cutting science areas
  - (**Carbon Cycle**; **Climate Variability**; **Atmospheric Comp**; **Disasters; Health & Air Quality**; **Atmospheric Composition Modeling and Analysis**; **Planetary Boundary Layer; Fires**)

- Further develop the High Performance Version of GEOS-Chem (GCHP), a global 3-D chemical transport model, to enable broad community access to high-resolution atmospheric chemistry modeling and chemical data assimilation
- Enhance community access to the cloud-enabled GCHP, a high-performance chemical transport model,
- Integrate the following technologies: high performance atmospheric chemistry model (GCHP); cloud computing; cloud pipelines; adaptive chemical solvers; smart load balancing; Modeling Analysis and Prediction Layer (MAPL);


GEOS-Chem as Offline and Online Chemical Module
any 3-D grid specified at run time
ESMF
ESMF
ESMF
ESMF
Advection
Mixing
Convection
Chemistry
(FlexChem):
dC/dt = P –L - D
Emissions
(HEMCO):
dC/dt = E
Off-line
GEOS-Chem CTM
GEOS-Chem chemical module
GEOS-Chem
User Community
10th International GEOS-Chem meeting in 2022 at WashU had 272 registrants from 103 institutions and 17 countries
ESTO
Earth Science Technology Office
5


GEOS-Chem as Offline and Online Chemical Module
any 3-D grid specified at run time
ESMF
ESMF
ESMF
Chemistry
(FlexChem):
dC/dt = P –L - D
Emissions
(HEMCO):
dC/dt = E
Dynamics,
chemical transport
GEOS-Chem chemical module
GEOS ESM with on-line
GEOS-Chem chemistry
GEOS ESM
ESTO
Earth Science Technology Office
6

GEOS-Chem as Offline and Online Chemical Module



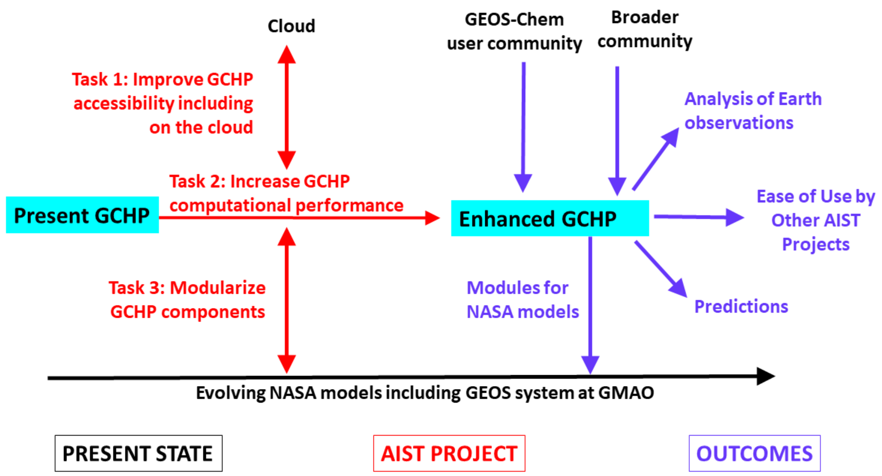
Summary of Initial Accomplishments

- **Updated to current version of MAPL and enable seamless updates** to benefit from future software engineering advances at GMAO for ESM coupling and data transfer, and to contribute MAPL enhancements produced by the GEOS-Chem community for atmospheric chemistry applications. MAPL developments at GMAO increased GCHP functionality and capabilities, such as the use of stretched-grid simulations.

- **Improved GCHP performance and portability** to promote usage across the large GEOS-Chem community and beyond. This effort included (a) parallelizing the model to remove current bottlenecks, thus enabling finer resolution simulations that better emulate the GEOS system, (b) improving the build system and using software containers to facilitate access and ease model configuration, and (c) developing support for a multi-node capability on the Amazon cloud.

- **Generated an operational cubed-sphere archive** of GEOS assimilated meteorological data for driving GCHP. This task increased accuracy in modeling transport and will provide a better foundation for collaboration between the GEOS-Chem community and GMAO.

Objectives of Augmentation



ESTO Earth Science Technology Office

9

Objectives of Augmentation

Further develop GCHP to support broad community access to high-resolution atmospheric chemistry modeling and chemical data assimilation, including by recently funded AIST (i.e., Huang and Keller)

Specific tasks:

1) Improve GCHP accessibility including on the cloud through a demonstration and documentation of capabilities, cataloging system to manage input data, and automated pipelines on the cloud;
2) Develop algorithms to increase computational performance through the use of adaptive chemical solvers, aerosol microphysics, and if time permits smart load balancing, to enable high-resolution global simulations with complex chemistry, and algorithms to increase the speed and efficiency of the models;
3) Modularize GCHP chemistry and thermodynamics solver to enable interoperability of these components within a next-generation ensemble prediction system for atmospheric composition (GEOS-CF) at NASA GMAO.

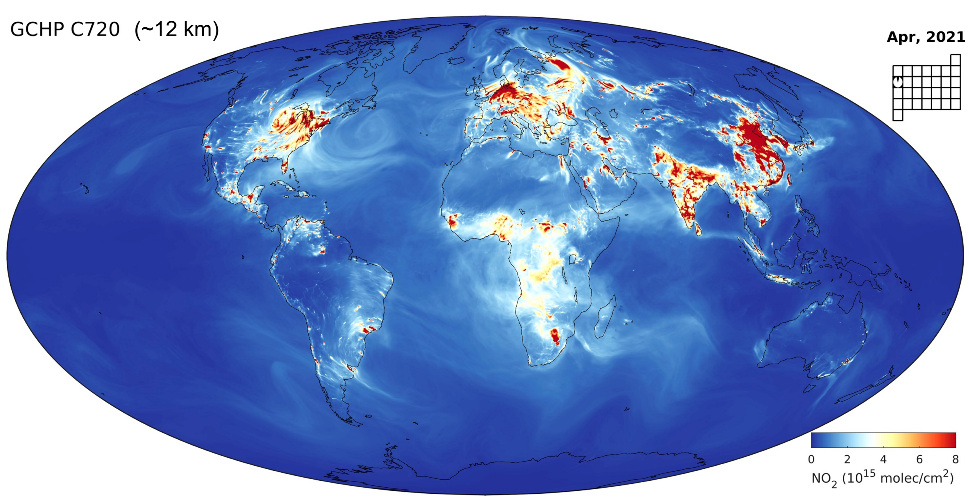
ESTO Earth Science Technology Office

10

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

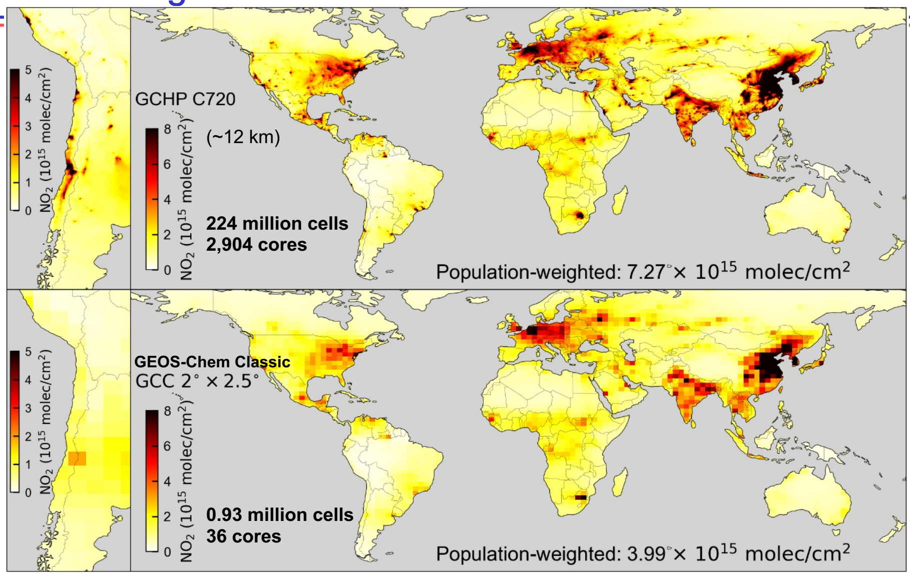
Task 1: Demonstration of GCHP



220 million cells, 2,904 cores

Dandan Zhang

Task 1: High Resolution Simulations with GCHP Reveal Large Differences in Ambient Concentrations



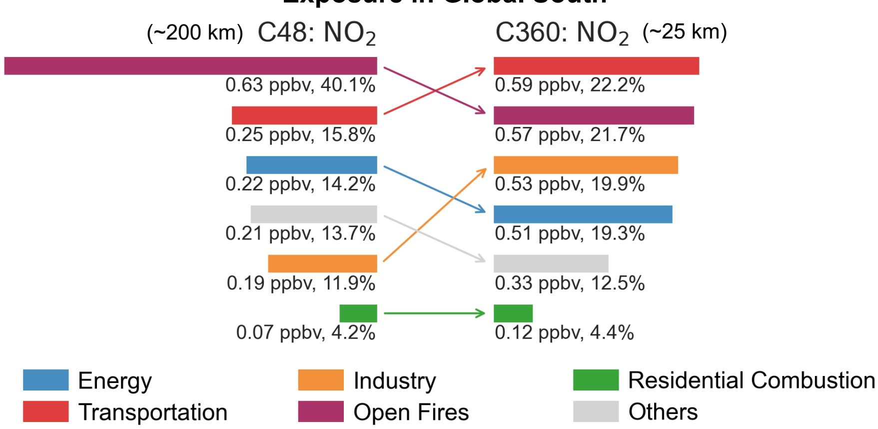
~240 times finer resolution

Martin et al., GMD, 2022

13

Task 1: Implications of Resolution: Altered Ranking of Sectoral Attribution

GCHP Sensitivity Simulations of Source Contributions to Population Exposure in Global South



Change in relative ranking has implications for air quality management strategies

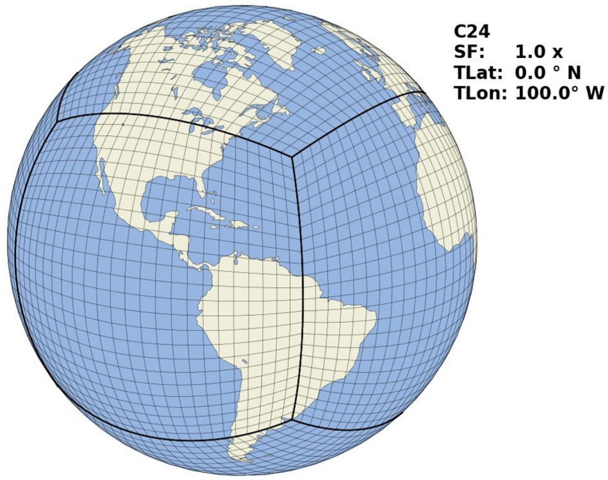
Zhang et al., 2023

14

Technical Development

Stretched-grid capability for targeted high-resolution simulations

- Transformation to the cube-sphere's grid-boxes
- Grid-boxes shrink over target region
- Grid-boxes expand on the opposite face
- No added computational effort



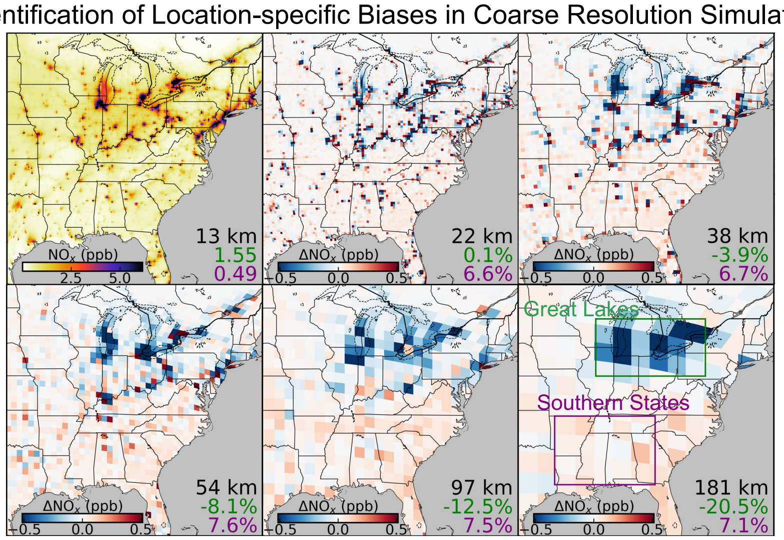
Bindle et al., GMD, 2021

ESTO Earth Science Technology Office

15

Task 1: Demonstration of GCHP Resolution Capabilities and Benefits

Identification of Location-specific Biases in Coarse Resolution Simulations



Regional mean values and biases inset

Li et al., ACP, 2023

ESTO Earth Science Technology Office

16

Task 1: Custom Cataloging System to Manage Input Data

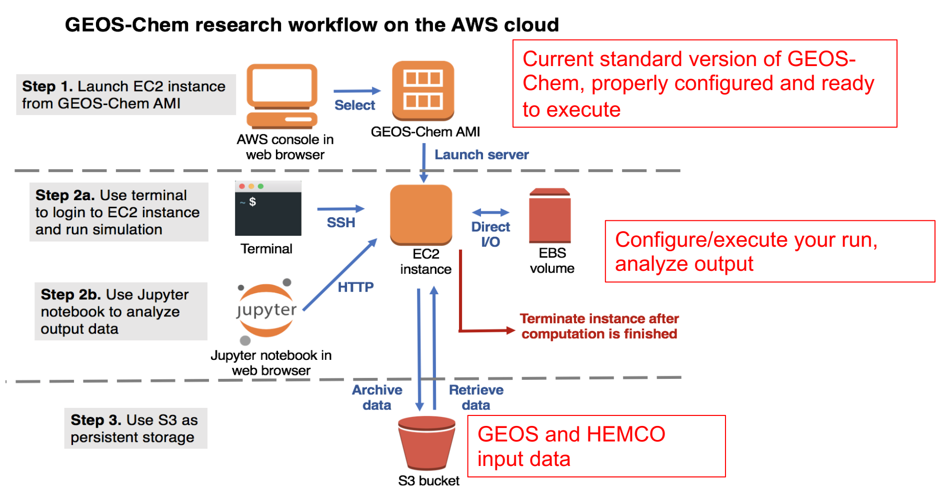
- Extensive input data for GCHP (> million files, > 100 TB)
- Created Bashdatacatalog to facilitate downloading and synchronizing local data collections with the GEOS-Chem online data repository
- Can query local data collections to identify missing files, or damaged files (not bitwise identical to remote files)
- Can format results into a) URL down list, b) Globus transfer list, c) rsync transfer list, or d) simple file list
- Dashboard available http://geoschemdata.wustl.edu/ExtData/Reports/CollectionSi

6/14/23, 12:00 PM geoschemdata.wustl.edu/ExtData/Reports/CollectionSizes.html

Input Data Collection	Total Size (Static)	Total Size (Temporal)	Asset Checksums	Temporal Asset Filters	Ignored Patterns
CHEM_INPUTS/FastJ_201204	0.000 GB	0.000 GB	.assets.md5		
CHEM_INPUTS/FAST_JX/v2020-02	0.000 GB	0.000 GB	.assets.md5		
CHEM_INPUTS/FAST_JX/v2021-10	0.000 GB	0.000 GB	.assets.md5		
CHEM_INPUTS/Linoz_200910	0.000 GB	0.000 GB	.assets.md5		
CHEM_INPUTS/Olson_Land_Map_201203	0.308 GB	0.000 GB	.assets.md5		
CHEM_INPUTS/UCX_201403	0.102 GB	0.000 GB	.assets.md5		.ignore.grep
GEOS_0.25x0.3125_AS/GEOS_FP	0.001 GB	3527.316 GB	.assets.md5	.temporal_assets.sed	

17

GCHP on the AWS cloud



Zhuang et al., JAMES 2020

18

Task 1: Progress on Cloud Pipelines: Application to Benchmarking

- Dashboard available at https://gc-c

6/14/23, 12:06 PM GEOS-Chem Testing Dashboard

GC Testing Dashboard

Registered Simulations

Filters ▼

Simulation ID	Date	Status	Code Url	Site	Description
gchp-c24-1Hr-14.2.0-alpha.14-3-g583bcfd	2023-06-05	SUCCESSFUL	583bcfd	AWS	1Hr gchp benchmark simulation using '583bcfd'
gcc-4x5-1Hr-14.2.0-alpha.14-3-gcb45840	2023-06-05	SUCCESSFUL	cb45840	AWS	1Hr gcc benchmark simulation using 'cb45840'
gchp-c24-1Hr-14.2.0-alpha.14-2-gc710de8	2023-06-05	SUCCESSFUL	c710de8	AWS	1Hr gchp benchmark simulation using 'c710de8'

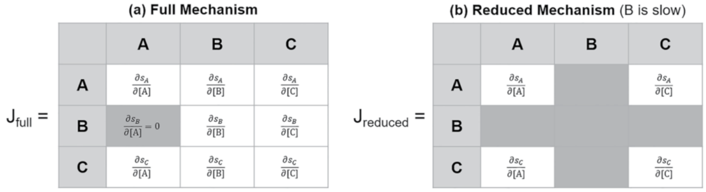
Difference Plots

ID	Date	Status	Site	Description
diff-gchp-c24-1Hr-14.2.0-alpha.14-1-g5f12cdb-gchp-c24-1Hr-14.2.0-alpha.14-3-g583bcfd	2023-06-05	SUCCESSFUL	AWS	1Hr Benchmark plot creation (ref: 'gchp-c24-1Hr-14.2.0-alpha.14-1-g5f12cdb'; dev:'gchp-c24-1Hr-14.2.0-alpha.14-3-g583bcfd')
diff-gchp-c24-1Hr-14.2.0-alpha.14-1-g5f12cdb-gchp-c24-1Hr-14.2.0-alpha.14-2-gc710de8	2023-06-05	SUCCESSFUL	AWS	1Hr Benchmark plot creation (ref: 'gchp-c24-1Hr-14.2.0-alpha.14-1-g5f12cdb'; dev:'gchp-c24-1Hr-14.2.0-alpha.14-2-gc710de8')

Tasks 2 & 3: Adaptive Chemical Solver (1 of 2)

Locally reduce chemical mechanism on the fly where full complexity not needed (slow species)

(R1) B + C → A + products (R2) A + C → products (R3) C + X → products



Sparse Jacobian:

```
NVAR:       3   LU_NONZERO:  8
LU_IROW:    1 1 1 2 2 3 3 3
LU_ICOL:    1 2 3 2 3 1 2 3
```

Sparse Jacobian:

```
NVAR:       3   LU_NONZERO:  8
LU_IROW:    1 1 1 2 2 3 3 3
LU_ICOL:    1 2 3 2 3 1 2 3

rNVAR:      2
cNONZERO:   4
JVS_MAP:    1 3 6 8
cLU_IROW:   1 1 2 2 - - - -
cLU_ICOL:   1 2 1 2 - - - -
```

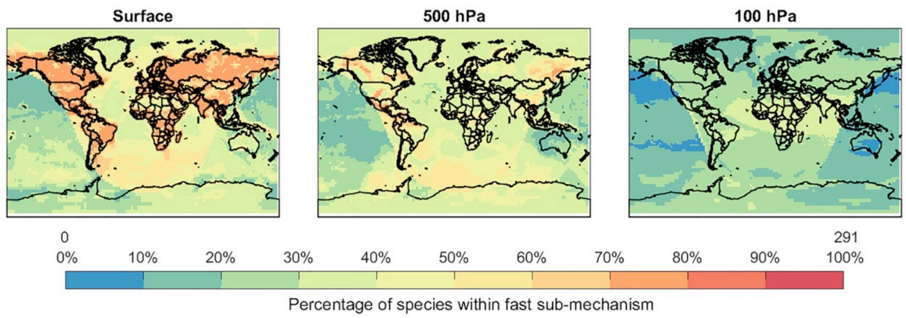
Reduced mechanism
With same data structure and memory space

Modular and released as KPP 3.0

Lin et al., JAMES, 2023

Task 2: Adaptive Chemical Solver (2 of 2)

Reduces cost of chemical calculations in free troposphere, stratosphere, and at night where chemistry less complex

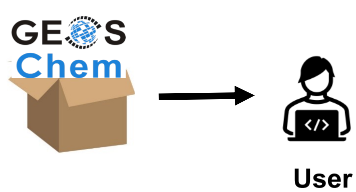


Benchmark tests achieved 30% speedup of chemical integration with errors < 1%

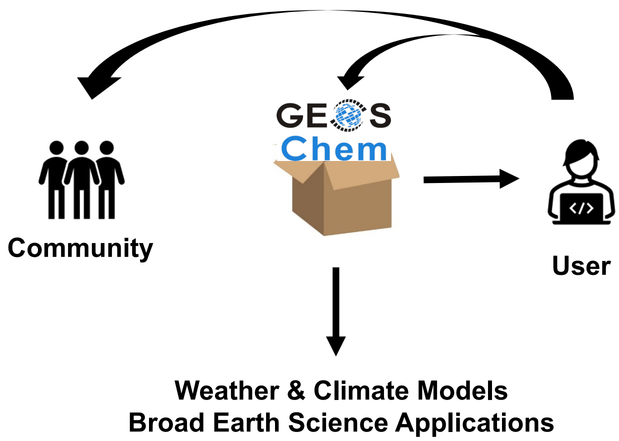
Full mechanism has 291 species and 913 reactions

Lin et al., JAMES, 2023

Accessibility to Contribute to GEOS-Chem Developments, Community, and Earth System Models



Accessibility to Contribute to GEOS-Chem Developments, Community, and Earth System Models



Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Current State

Task 1: Completed demonstration of capabilities and cataloging system for input data. Continuing to develop automated pipelines on the cloud;

Task 2: Completed adaptive chemical solver. Developing aerosol microphysics capability and conducting studies of smart load balancing;

Task 3: Completed modularization of chemical solver. Conducting work to modularize thermodynamics solver.

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Actual or Potential Transfers and Infusions

- Technology transfer = transferring technology via licensing, open-sourcing, or public domain release (including through a NASA Technology Transfer Program)
- **GCHP is publicly available at www.geos-chem.org with documentation at gchp.readthedocs.io. The latest GCHP version (14.1.1) is available at zenodo.org/record/7696683.**

- Infusion = adoption of technology by an individual organization or project
- **Known infusions: Harvard, MIT, Washington University, CU Boulder, NASA Jet Propulsion Laboratory, University of York, Laboratoire d'Aerologie, University of Wollongong, University of Toronto, Colorado State University, Centre for Research on Energy and Clean Air, Peking University, Southern University of Science and Technology, Tsinghua University**

- **Potential infusions: the GEOS-Chem community, air quality managers, and beyond**

- Knowledge transfer = transfer of ideas or knowledge generated and included in an AIST funded technology: **AQ ACF, GEOS-CF**

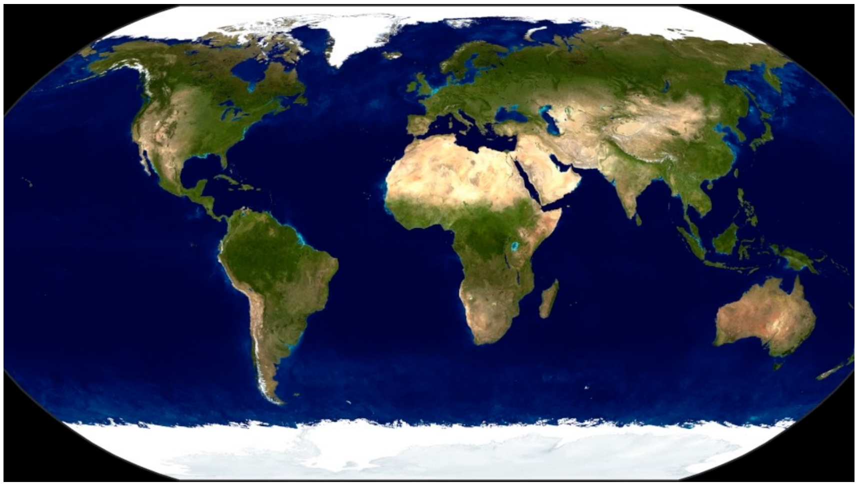
Actual or Potential Transitions and Collaborations

- Transition = when a technology continues to evolve or develop within a new program (NASA or non-NASA) including transition to a new domain application such as from the space domain to the medical domain.
- **Current Transition: GMAO**
- **Potential Transition:**
  - **NASA Carbon Monitoring System Flux Project (CMS-Flux)**
  - **NCAR Community Earth System Model (CESM)**
  - **NOAA Unified Forecast System (UFS)**
  - **NASA TEMPO satellite mission**
  - **Environmental health domain**

- Provide a summary of actual or potential collaborations
  - Including NASA or non-NASA collaborations
  - **Atmospheric chemistry community**
  - **Satellite remote sensing community**
  - **Environmental health community**

Potential to Infer CO_2 Fluxes From Satellite Observations

GEOS-Chem High Performance, C360



Kevin Bowman (JPL) and Junjie Liu (Arizona State University)

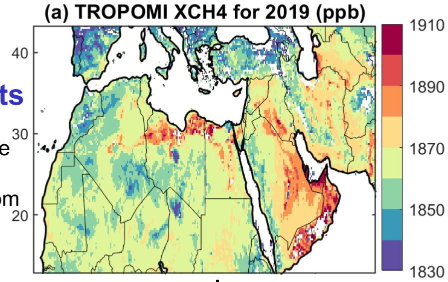
ESTO Earth Science Technology Office

29

Potential to Extend to Methane Fluxes

Using TROPOMI to evaluate UNFCCC national emission reports

- Optimize emissions for 23 countries in Middle East and N. Africa at 25-km resolution
- Now tiling the world to quantify emissions from all countries



Inversion using GEOS-Chem

Prior ≡ UNFCCC

Country		Oil	Gas	Livestock	Coal	Waste	Rice	Other	Anthropogenic total ($Tg\ a^{-1}$)	Sensitivity to observations[b]
Algeria	Posterior	0.08	2.0	0.43	0	0.95	0	0.01	3.5 (2.4-4.4)	0.84
	Prior	0.04	1.0	0.29	0	0.65	0	0.01	2.0	
Bahrain	Posterior	0.14	0.07	0	0	0.19	0	0.02	0.42 (0.39-0.43)	0.91
	Prior	0.17	0.03	0	0	0.06	0	0.01	0.27	
Egypt	Posterior	0.35	0.15	0.78	0	1.6	0.35	0.12	3.4 (2.5-4.0)	0.96
	Prior	0.24	0.07	0.41	0	0.71	0.15	0.06	1.7	
Iran	Posterior	0.78	1.0	1.20	0.04	1.7	0.22	0.32	5.3 (5.0-5.5)	0.97
	Prior	2.6	0.52	0.65	0.02	0.81	0.13	0.14	4.9	
Iraq[c]	Posterior	1.2	0.04	0.17	0	0.73	0.02	0.05	2.2 (1.8-3.1)	0.98
	Prior	2.9	0.03	0.14	0	0.54	0.01	0.04	3.7	

Etc. (all 23 countries)

Chen, Jacob, et al., ACP, 2023

30

ESTO Earth Science Technology Office

30

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Publications (1 of 2)

Journal / Conference Papers

Zhuang, J., D.J. Jacob, H. Lin, E.W. Lundgren, R.M. Yantosca, J. Flo Gaya, M.P. Sulprizio, S.D. Eastham, and K. Jorissen, Enabling high-performance cloud computing for Earth science modeling on over a thousand cores: application to the GEOS-Chem atmospheric chemistry model, Journal of Advances in Modeling Earth Systems, doi: 10.1029/2020MS002064, 2020.

Bindle, L., Martin, R. V., Cooper, M. J., Lundgren, E. W., Eastham, S. D., Auer, B. M., Clune, T. L., Weng, H., Lin, J., Murray, L. T., Meng, J., Keller, C. A., Putman, W. M., Pawson, S., and Jacob, D. J., Grid-stretching capability for the GEOS-Chem 13.0.0 atmospheric chemistry model, Geosci. Model Dev., 14, 5977–5997, https://doi.org/10.5194/gmd-14-5977-2021, 2021.

Meng, J., Martin, R. V., Ginoux, P., Hammer, M. S., Sulprizio, M. P., Ridley, D. A. and van Donkelaar, A., Grid-independent high-resolution dust emissions (v1.0) for chemical transport models: application to GEOS-Chem (12.5.0)., Geosci. Model Dev., doi:10.5194/gmd-14-4249-2021, 2021.

Martin, R. V., Eastham, S. D., Bindle, L., Lundgren, E. W., Clune, T. L., Keller, C. A., Downs, W., Zhang, D., Lucchesi, R. A., Sulprizio, M. P., Yantosca, R. M., Li, Y., Estrada, L., Putman, W. M., Auer, B. M., Trayanov, A. L., Pawson, S., and Jacob, D. J.: Improved advection, resolution, performance, and community access in the new generation (version 13) of the high-performance GEOS-Chem global atmospheric chemistry model (GCHP), Geosci. Model Dev., 15, 8731–8748, https://doi.org/10.5194/gmd-15-8731-2022, 2022.

Publications (2 of 2)

Journal / Conference Papers

Li, C., Martin, R. V., Cohen, R. C., Bindle, L., Zhang, D., Chatterjee, D., Weng, H., & Lin, J.. Variable effects of spatial resolution on modeling of Nitrogen Oxides. Atmospheric Chemistry and Physics, 23(5), 3031–3049. https://doi.org/10.5194/acp-23-3031-2023, 2023.

Lin, H., M.S. Long, R. Sander, A. Sandu, R.M. Yantosca, L.A. Estrada, L. Shen, and D.J. Jacob, An adaptive auto-reduction solver for speeding up integration of chemical kinetics in atmospheric chemistry models: implementation and evaluation in the Kinetic Pre-Processor (KPP) version 3.0.0, JAMES, 15, e2022MS003293. https://doi.org/10.1029/2022MS003293, 2023.

Zhang, D., Martin, R. V., Bindle, L., Li, C., Eastham, S. D., van Donkelaar, A., Gallardo, L.: Advances in Simulating the Global Spatial Heterogeneity of Air Quality and Source Sector Contributions: Insights into the Global South. Environmental Science & Technology. https://doi.org/10.1021/acs.est.2c07253, 2023.

Presentations (1 of 2)

Eastham, S. D., Chossière, G., Speth, R. L., & Barrett, S.R.H. The role of aviation and intercontinental transport in local air quality. American Geoscientists Union (AGU) Annual Fall Meeting, 2019.

Eastham, S. D., Monier, E., Rothenberg, D., & Selin, N. Time of emergence for the influence of climate change on surface ozone. American Meteorological Society (AMS) Annual Meeting, 2020.

Jacob, D.J. and R.V. Martin, GEOS-Chem model overview, Joint keynote presentation, 1st GEOS-Chem Europe Meeting, 1 September 2020.

Martin, R.V., Progressing from Global to Urban Scales for Air Quality Applications, Earth Science Information Partners Virtual Meeting, 15 July 2020.

Martin, R.V., Advancing Understanding of Air Quality from Global to Urban Scales, Frontiers of Atmospheric Science, American Geophysical Union Virtual Conference, December 2020.

Martin, R.V., Fine Particulate Matter: Interpreting Satellite Observations to Advance Understanding for Health Applications, ASEEP Lecture, American Association for Aerosol Research plenary presentation, Albuquerque, 21 Oct 2021.

Presentations (2 of 2)

- Martin, R.V., Overview of GEOS-Chem, 10th International GEOS-Chem Meeting, Washington University, 7 Jun 2022.
- Bindle, L. and Lundgren, E., Working with the high-performance GEOS-Chem (GCHP), 10th International GEOS-Chem Meeting, Washington University, 6 Jun 2022.
- Zhang, D., Advances in simulating the spatial heterogeneity of air quality and source contributions using GCHP, 10th International GEOS-Chem Meeting, Washington University, 6 Jun 2022.

Acronyms

List of Acronyms

- ESMF Earth System Modeling Framework
- GCHP GEOS-Chem High Performance
- GEOS Goddard Earth Observation System
- GMAO Global Modeling and Assimilation Office
- HEMCO Harvard-NASA EMission Component
- KPP Kinetic PreProcessor
- MAPL Modeling Analysis and Prediction Layer
- TRL Technology Readiness Level

A Framework for Global Cloud-Resolving OSSEs

Thomas Clune(PI, NASA GSFC)
Arlindo da Silva (Science-PI, NASA GSFC)

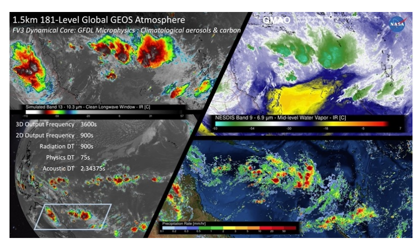
AIST-18/21-0063 Annual Technical Review
June 23, 2023

Team listing: W. Putman (GSFC), R. Todling (GSFC),
V. Buchard (GSFC & ESSIC), D. Posselt (JPL), & D. Duffy (GSFC)

A Framework for Global Cloud-Resolving OSSEs

PI: Thomas Clune, NASA GSFC

To design, implement, and deploy an efficient framework for efficiently performing high spatial and temporal resolution global cloud-resolving Observing System Simulation Experiments (OSSEs). Among the many applications of global cloud resolving OSSEs is the development of missions comprising NASA's Earth System Observatory (ESO), from the design of space mission architectures, to the exploration of the possibilities that the new observing system offers to science and applications, well before launch.



Example GEOS Non-Hydrostatic DYAMOND Phase-II Simulations. The aim is to simulate the most important scales of the full three-dimensional fluid dynamics of the atmospheric circulation

- Achieve high temporal resolution O(~1 minute) by leveraging the identically configured model to generate intermediate model states on demand from a relatively thin set of reference nature run checkpoints.
- Define a simple API for custom geometry products and implement 4 concrete instances. Allow for custom data formats including JEDI IODA.
- Demonstrate framework by using custom geometry products with a suite of L1 simulators.

Co-Is/Partners: A. da Silva, W. Putman, R. Todling, V. Buchard, D. Posselt (JPL), & D. Duffy

• Requirements analysis and design	01/23
• Reference Nature Runs Completed	~~05/23~~ 09/23
• All samplers implemented	07/23
• All L2 Simulations Completed	~~09/23~~ 11/23
• All L1 Simulations Completed	07/24

TRL_{in} = 3 $TRL_{current}$ = 3

08/22 AIST-21-0063 2

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Project Objectives

Science and Technology Objectives

- Enable global storm resolving Observing System Simulation Experiments (OSSEs)
- Enable simulation of past, present and future observing systems from global storm resolving Nature Runs.
- Address computational challenges that prevent existing technologies from scaling to spatial resolutions that will be needed by end of the decade.

Application Objectives

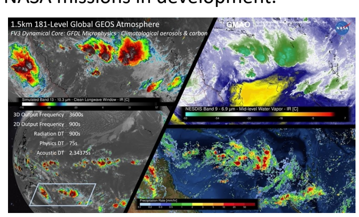
- Leverage and engage the *Joint Effort for Data Assimilation Integration* (JEDI) community by providing simulation of NWP observing system including all-sky radiances, accounting for the effect of aerosols
- Enable simulation of PBL related observables, engaging the *Decadal Survey Incubator* (DSI) PBL community
- Engage the Earth System Observatory (DSO) missions in development (AOS, SBG) by providing atmospheric Nature Runs for realistic observing system simulations enabling algorithm development, science and application demonstrations well before launch.

Benefits to Earth Science

- Bridges the gap between advanced earth system modeling and realistic simulation of future observing systems
- Early development of new concepts in model-data fusion enabling science investigations aimed at elucidating physical processes and their representation in models.

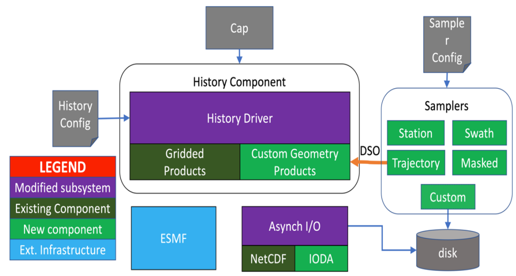
Relationship to ES Digital Twins

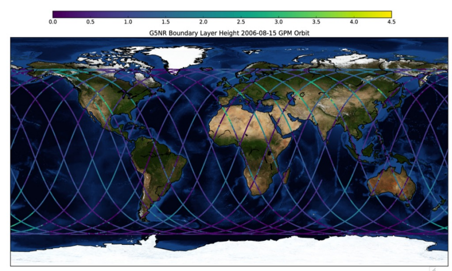
- Develops a framework for producing advanced global storm Nature Runs that are **Digital Replicas** of the Earth System
- Provide a framework for observing system simulations and tools to conduct *what-if* investigations and trade studies for the benefit of NASA missions in development.



Technology Approach

- Extend parallel I/O capabilities of the GEOS model by including on-line sampling of *Geophysical Variables* at observation locations
  - ground stations, aircraft trajectories, satellite swaths
- Two-phase workflow for generating Nature Runs
  1) Model is spun up in a run with limited output except for frequent checkpoints and browse products
  2) Model is run from spun up checkpoints with output sampled at user specified observation locations
- Definition of a flexible API for custom geo-location samplers, allowing maximum flexibility





Presentation Contents

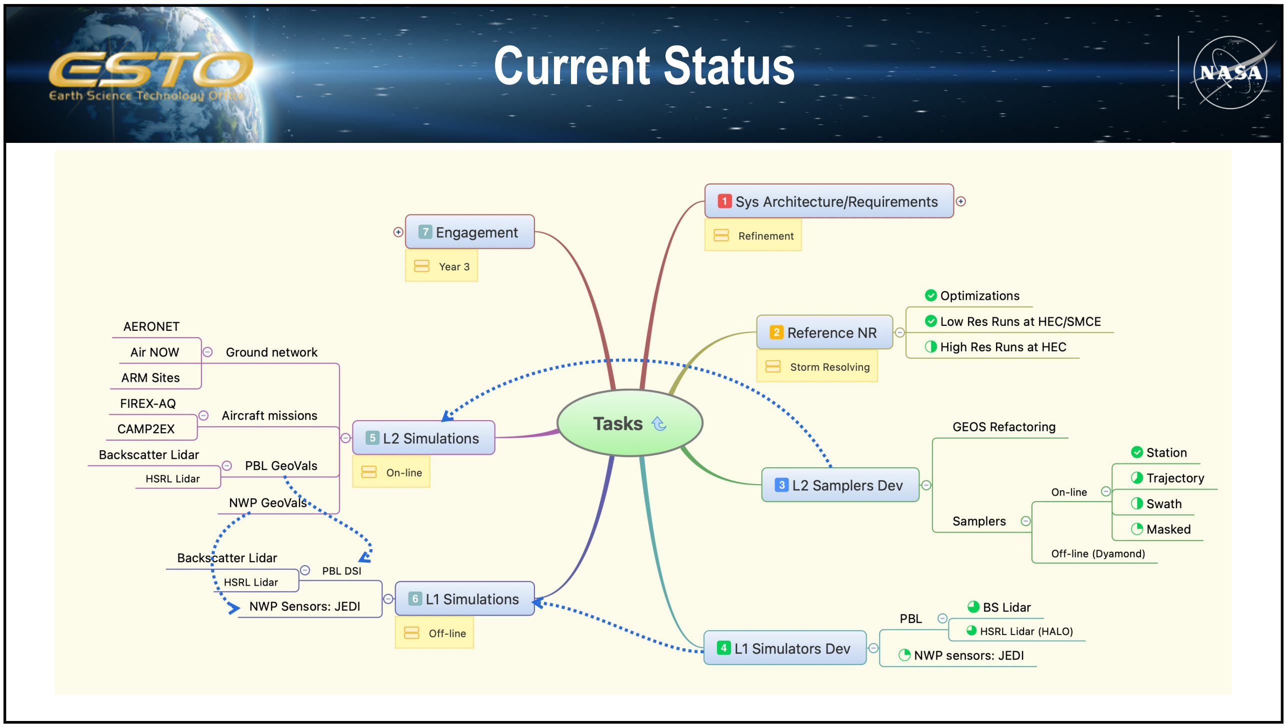
✓Background and Objectives

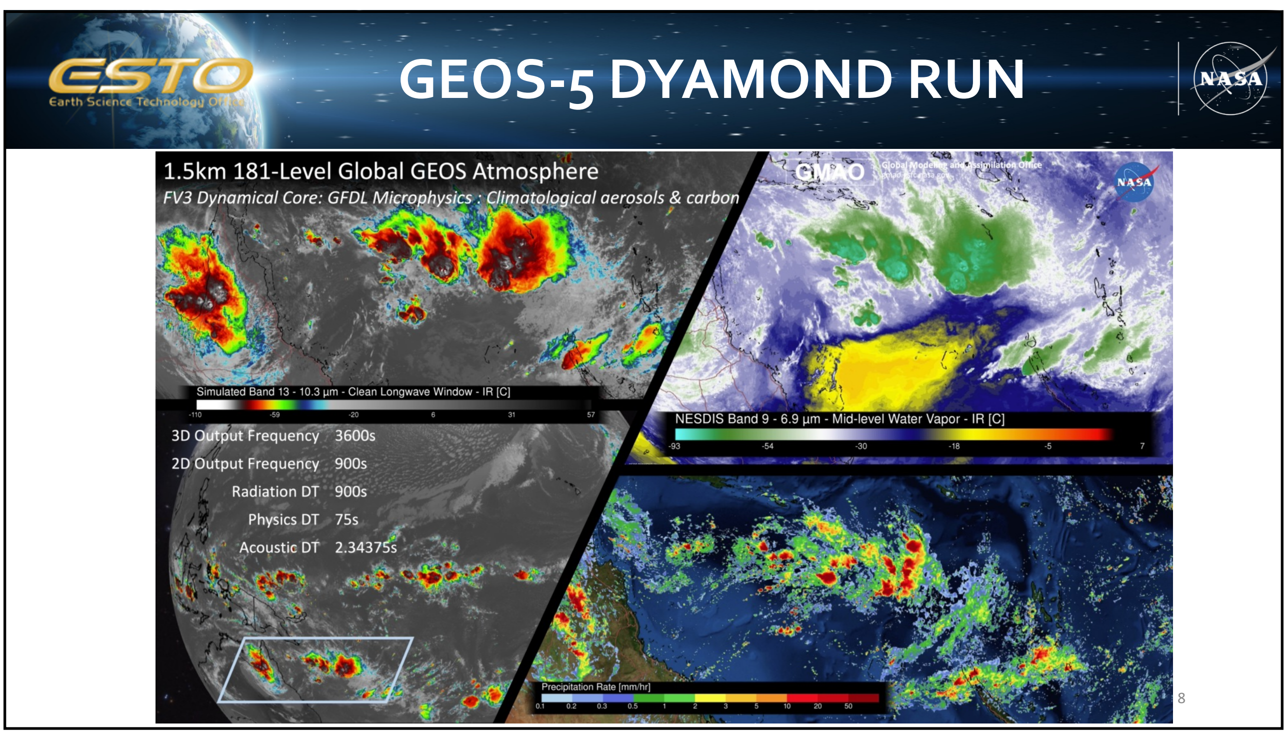
○Technical and Science Advancements

○Summary of Accomplishments and Future Plans

○Infusions and Collaborations

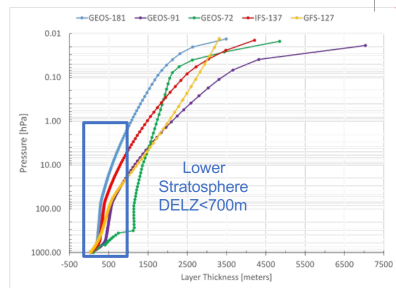
○Publications - List of Acronyms

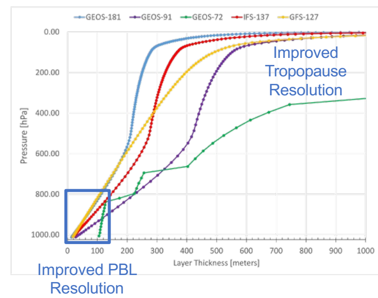

ESTO
Earth Science Technology Office
Current Status
NASA
1 Sys Architecture/Requirements
Refinement
7 Engagement
Year 3
Optimizations
Low Res Runs at HEC/SMCE
High Res Runs at HEC
2 Reference NR
Storm Resolving
AERONET
Air NOW
Ground network
ARM Sites
FIREX-AQ
Aircraft missions
CAMP2EX
Backscatter Lidar
PBL GeoVals
HSRL Lidar
NWP GeoVals
5 L2 Simulations
On-line
Tasks
GEOS Refactoring
Station
Trajectory
Swath
Masked
On-line
Samplers
Off-line (Dyamond)
3 L2 Samplers Dev
Backscatter Lidar
PBL DSI
HSRL Lidar
NWP Sensors: JEDI
6 L1 Simulations
Off-line
PBL
BS Lidar
HSRL Lidar (HALO)
NWP sensors: JEDI
4 L1 Simulators Dev


ESTO
Earth Science Technology Office
GEOS-5 DYAMOND RUN
NASA
1.5km 181-Level Global GEOS Atmosphere
FV3 Dynamical Core: GFDL Microphysics : Climatological aerosols & carbon
GMAO
Simulated Band 13 - 10.3 µm - Clean Longwave Window - IR [C]
NESDIS Band 9 - 6.9 µm - Mid-level Water Vapor - IR [C]
3D Output Frequency 3600s
2D Output Frequency 900s
Radiation DT 900s
Physics DT 75s
Acoustic DT 2.34375s
Precipitation Rate [mm/hr]
8

Earth System Model: GEOS-6

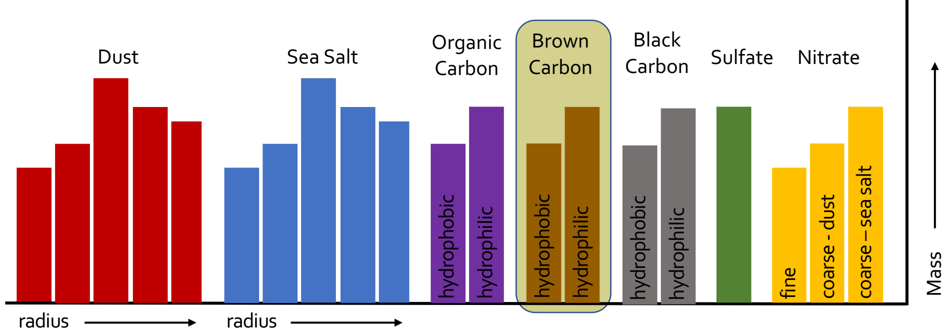
- ✓ **GEOS-6 'Krok' v11 Release**
  - support for 72L, 91L, 137L and 181L and stretched grid options
- ✓ Non-Hydrostatic FV3/TRB/GWD/MST/UW [reduce vertical resolution sensitivity]
- ✓ Refactored GEOS moist physics [two primary microphysics options]:
  - GFDL-Microphysics [6-phase, single moment, numerous NASA modifications]
  - Morrison-Gettleman-Barahona 2-moment cloud microphysics [MG3]
  - GF2020 [Grell-Freitas Deep convective updates and new tunings]
  - UW Shallow Convection [improved connections with Turbulence/PBL]
  - CCPP option for running with GFSv16 physics
- ✓ GWD tuning for improved QBO and NWP skill
  - NCAR new GWD [background and mountain ridge orographic schemes]
  - BKG: Richter et al, 2010: https://doi.org/10.1175/2009JAS3112.1
  - ORO: Based on Scinocca and McFarlane, 2000: https://doi.org/10.1002/qj.49712656802
- ✓ Land model [new MODIS snow albedo treatment]
- ✓ EDMF-SHOC [unified boundary layer scheme]
  - Completed but disabled by default
- ✓ Includes option to run RRTMGP LW+SW
- ✓ GOCART-2G [Improved efficiency and reduced code complexity]





Science Changes Bundled with GOCART-2G

- Separation of organic aerosol into "white" (anthropogenic) and "brown" (biomass burning) components with distinct optical properties
- Increase OA:OC ratio in line with recent airborne measurements
- Inclusion of an ACHEM-driven **SOA scheme** for anthropogenic and biomass burning sources
- Inclusion of a **HEMCO/MEGAN-driven biogenic** SOA scheme
- Introduction of "point wise" source emissions for pyroCb inputs
- Update anthropogenic emissions to downscaled-**CEDS emission** inventory and input oxidant fields to MERRA-2 GMI (valid range of both is 1980 - 2019; padding outside years with endpoints



Pete

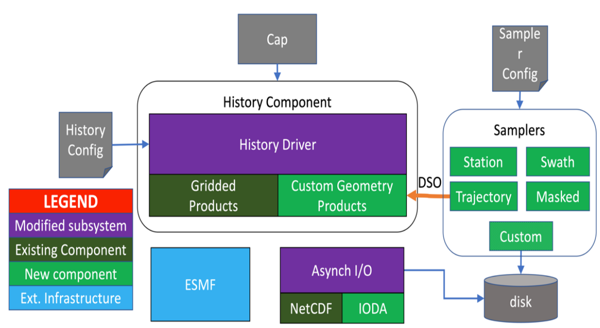
Toward a Reference Nature Run

- Refactored the GEOS Moist Physics component
  - Reduced memory footprint by eliminating extra copies of tracers
- Addressed other issues at high processor count
  - Eliminated global array allocation on all sub-domains
  - Addressed corner case involving coarse I/O at high processor count (work starvation)
- Demonstrated 3 km global runs on 25K and 50K cores
  - Throughput is ~15 days/day on 50K cores with 75s physics timestep
- With those fixes, we care confident that 1.5 km run now fits (an runs) in 50K core footprint
  - Estimate ~2-3 days/day for 1.5 km with 30s-45s physics timestep
  - Reservation on Aitken for next week
- (A lower resolution run has also been demonstrated on AWS on a pcluster.)

2023 AIST Annual Review 11

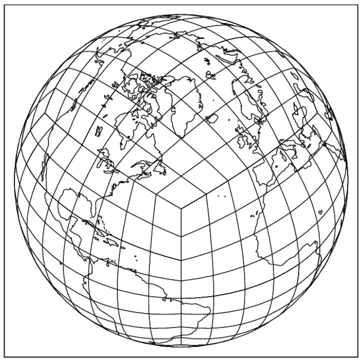
Extensions of GEOS History Component

- ***History*** is a flexible component enabling all GEOS product suites
  - Choice of any variable exported by any component
  - New variables defined by a powerful expression parser
  - Runtime specified grids and resolutions
  - Arbitrary time frequency
  - Instantaneous or time average output
- This project extends ***History*** by providing output on new measurement-oriented geometries
  - Stations, trajectories, swath
  - Retains all other flexibility



2023 AIST Annual Review 12

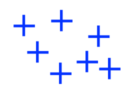
Foundation: ESMF Regridding



GEOS Atmospheric Grid
(ESMF Cubed-sphere Grid)

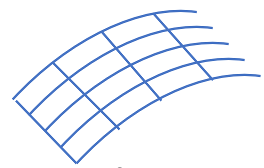
ESMF Regrid

Various regrid options:
- Bilinear
- Conservative
- Masking
- Etc.



ESMF
LocStream

- Station
- Trajectory

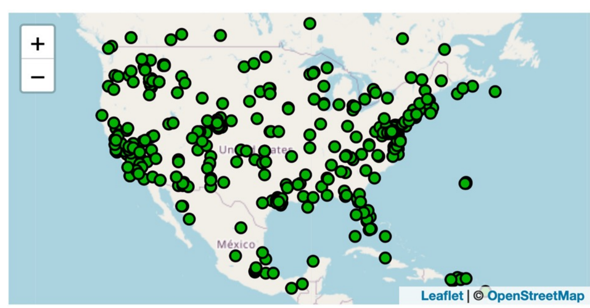


ESMF
Logically rectangular Grid

- Swath

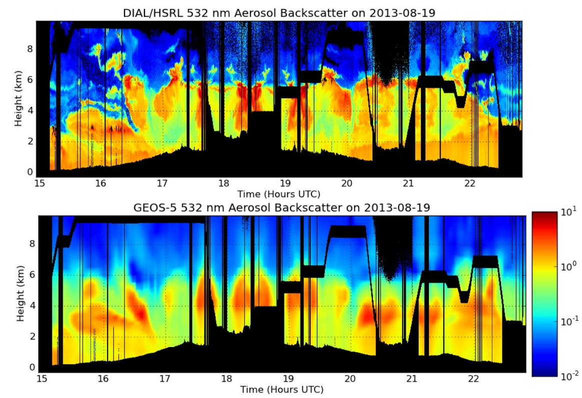
Station Sampler

- Produce time series output at list of fixed geolocations
- Implementation
  - Geolocations represented as *fixed* LocStream
  - ESMF Regrid weights computed just once at beginning of execution



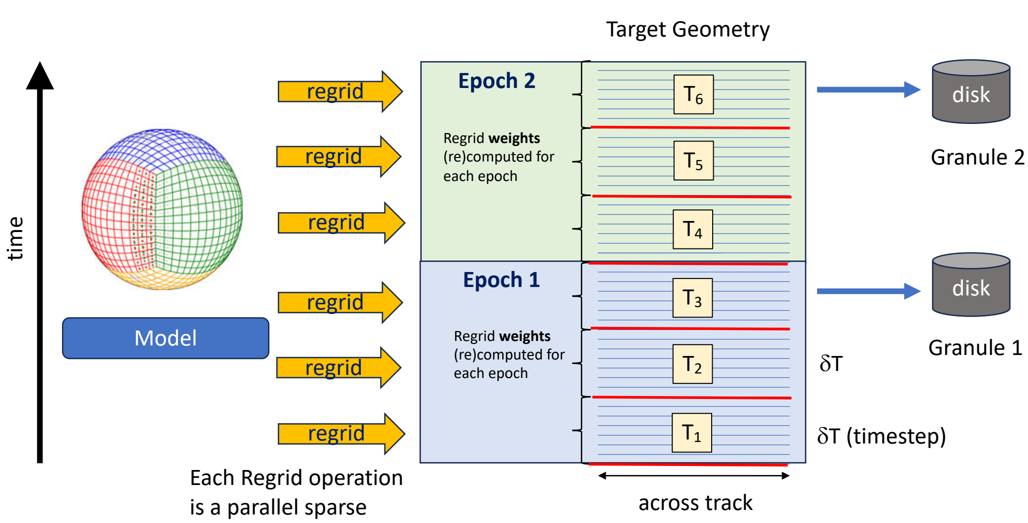
Trajectory Sampler

- In its simplest form, a trajectory is aircraft or satellite track indexed by time
  - Example: lidar curtain
- More generally, it is a sequence of triplets (time,longitude,latitude) which can be sorted by time
  - Example: NWP data after cloud screening and quality control
- Approach:
  - Nearest neighbor time interpolation within model time step
  - ESMF regrid capability used to go from cubed-sphere to location streams



GEOS-5 Backscatter compared to DIAL/HSRL Lidar on August 19th, 2013 during SEAC4RS.

Swath sampler



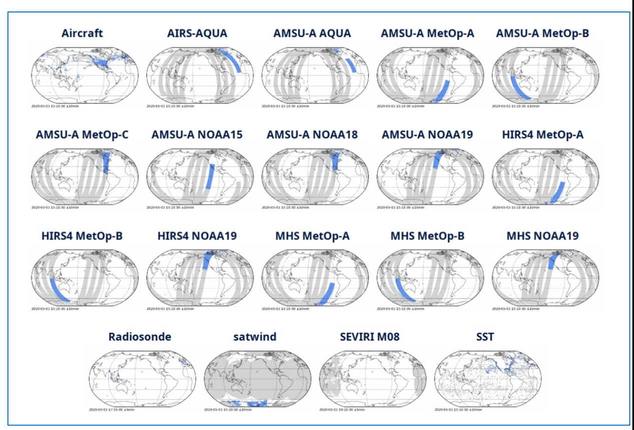
Each Regrid operation is a parallel sparse Matrix vecor multiply

Masked Sampler

- Basic concept:
  - Output variables on full model grid
  - Mask out unobserved gridcells
  - Rely on compression to reduce data volume & reduce I/O time
- Useful when swath resolution >> model resolution
  - Regridding to finer grid is unnecessarily expensive
- Desired: include halo (ghost) cells in output to support offline gradients
- Challenges
  - Cannot combine multiple timesteps into one file (obs can overlap)
  - Computing mask halo is a bit more convoluted than one might expect

Observation Simulators

- **Lidar and PBL**
  - Lidar observables (backscatter, extinction and total attenuated backscatter) have been implemented natively in GEOS
  - Sampling these along lidar trajectory will provide the relevant L1 simulations
  - While PBL Height is available as native GEOS diagnostic, and a lidar derived estimate (from extinction and backscatter profiles) are yet to be developed
- An in-depth analysis has been conducted and strategy has been devised for simulation the conventional and satellite **NWP observing system**.
- Work on L1/L2 simulations will start on next quarter



All operational NWP instruments used at NOAA/NWS/EMC for global GFS forecasts have been implemented in JEDI's Unified Forward Operator (UFO) component.

https://www.jcsda.org/jcsda-project-jedi

Presentation Contents

- ✓Background and Objectives
- ✓Technical and Science Advancements
- ○Summary of Accomplishments and Future Plans
- ○Infusions and Collaborations
- ○Publications - List of Acronyms

Summary of Accomplishments

- Have identified several scalability concerns/bugs
  - Have demonstrated 3km resolution on 50k cores with desired GEOS config
  - Optimistically, no new problems will be revealed at 1.5k / 50k cores.
- Samplers
  - Implementation of **station** and **swath** samplers is complete
  - **Trajectory** is near completion – optimization for large trajectories
  - Initial design of **masked** sampler complete
- Created capability for testing using offline data for testing L2 simulators.
  - A device for sampling and previously produced NR

Future Plans

- Once 1.5 km configuration has been verified, reproduce complete workflow with low resolution NR:
  - Produce low resolution reference NR
  - Sample geophysical variables at all geometries, including complete NWP observing system
  - Complete development of L1 simulators
  - Produce Level 1 & 2 simulations based on low-res NR
- Start production runs for high-res NR (1-2 months of wall time!)
- Once high-res NR is complete, exercise/optimize full workflow
  - Demonstrate ability to retrieve high spatiotemporal samples
  - Profile/optimize system, ensuring scalability
  - Produce L1/L2 simulations based on high-resolution NR

Presentation Contents

✓Background and Objectives

✓Technical and Science Advancements

✓Summary of Accomplishments and Future Plans

○Infusions and Collaborations

○Publications - List of Acronyms

Infusions and Collaborations

- **Infusion:** AOS
  - Global storm resolving nature runs will be provided to the Atmospheric Observing System (AOS) for simulation of their observing system and design of their science data systems.

- **Collaborations**
  - Other NASA centers (JPL, ARC): through close with AOS project
  - NCAR (ESMF core team): optimization of ESMF regridding with custom geometries
  - Collaboration with the 2017 Decadal Survey Incubator (DSI) PBL including multiple investigators from GSFC, JPL, and LaRC

Presentation Contents

✓Background and Objectives

✓Technical and Science Advancements

✓Summary of Accomplishments and Future Plans

✓Actual or Potential Infusions and Collaborations

○Publications - List of Acronyms

Publications

- None yet

List of Acronyms

• AOS	Atmosphere Observing System
• C5760	Cubed-sphere with 5760 gridcells per edge. ~1.5 km global resolution
• DSI	Decadal Survey Incubator
• GEOS	Goddard Earth Observing System Earth System Model
• ESMF	Earth System Modeling Framework (ESTO funded)
• NR	Nature Run
• OSSE	Observing System Simulation Experiment
• TLE	Two-line element orbital specification

GEOS Visualization And Lagrangian dynamics Immersive eXtended Reality Tool (VALIXR) for Scientific Discovery

Thomas G. Grubb (PI, GSFC)
Tom Clune (Co-I, GSFC)
Leslie Lait (Co-I, GSFC)
Steve Guimond (Co-I, UMBC)
Matthias Zwicker (Co-I, UMD)
Ruth West (Co-I, UNT)

AIST-21-0052 Annual Technical Review
6/23/2023

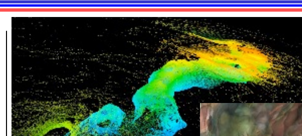
GEOS Visualization And Lagrangian dynamics Immersive eXtended Reality Tool (VALIXR) for Scientific Discovery

PI: Thomas G. Grubb / GSFC

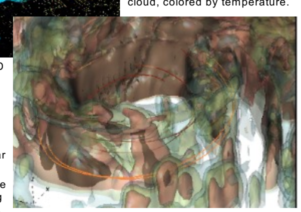
Objective

Develop a scientific exploration and analysis mixed reality (XR) tool with integrated Lagrangian Dynamics (LD) for the Goddard Earth Observing System (GEOS) numerical weather prediction model to help scientists identify, track, and understand the evolution of Earth Science phenomena. VALIXR will:

- Enhances GEOS to calculate Lagrangian trajectories of specific Earth Science phenomena and output budget terms (e.g., momentum) and parcel attributes (e.g., temperature) that describe their dynamics
- Enhance the NASA open source XR tool, the Mixed Reality Exploration Toolkit (MRET), to visualize and animate GEOS fields and new LD trajectories
- Enable scientists to identify, track and understand various ES phenomena in a Lagrangian reference frame with budget terms to describe their dynamics



Top: Viz of Point Clouds in VR. A similar viz of GEOS model data will be created; imagine an animated smoke plume point cloud, colored by temperature.



Right: Current, state of the art, 2D viz using 3D isosurfaces of convective clouds (w/opacity scaling to see inside the clouds). Colored lines represent the trajectories of air parcels moving through the storm, which is similar to the LD tool envisioned in VALIXR. It is difficult to view these lines and convective clouds using standard visualization techniques.

Approach

1. Create an open-source point cloud visualization component for MRET
2. Enhance MRET with LD tools for the initialization, finalization and visualization of trajectories with associated dynamic budget terms, and to output a parameter package for use in GEOS model runs
3. Integrate GTraj into GEOS to calculate Lagrangian trajectories of specific Earth Science phenomena and output budget terms that describe their dynamics
4. Apply VALIXR to Earth Science phenomena (deep moist convection and wildfires)

Key Milestones

• Requirements, preliminary design, prototyping	12/22
• VALIXR B1 (PC beta, LD, GTraj, TRL 4)	07/23
• VALIXR B2, Use Cases, Human Factors (TRL 5)	01/24
• VALIXR Final, Full Use Cases (TRL 6)	07/24

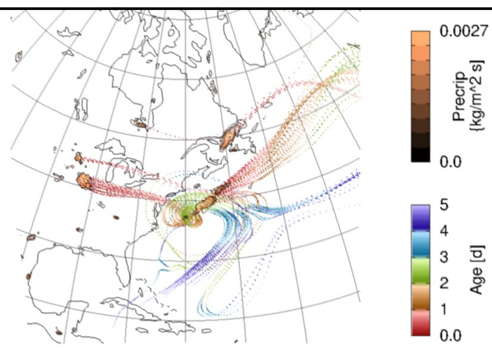
Co-Is/Partners: Tom Clune, Leslie Lait, GSFC, Steve Guimond, Don Engel, UMBC; Matthias Zwicker, Roger Eastman, UMD; Ruth West, UNT

TRL_{in} = 3 $TRL_{current}$ = 3

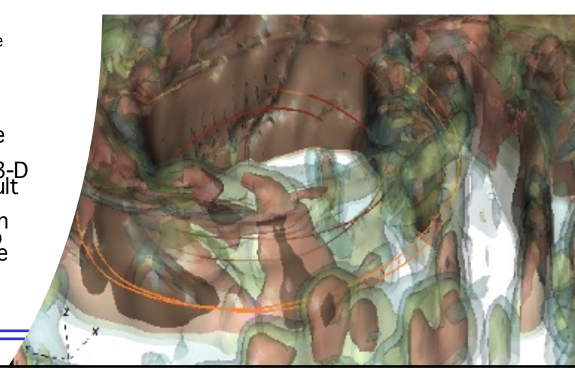
Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

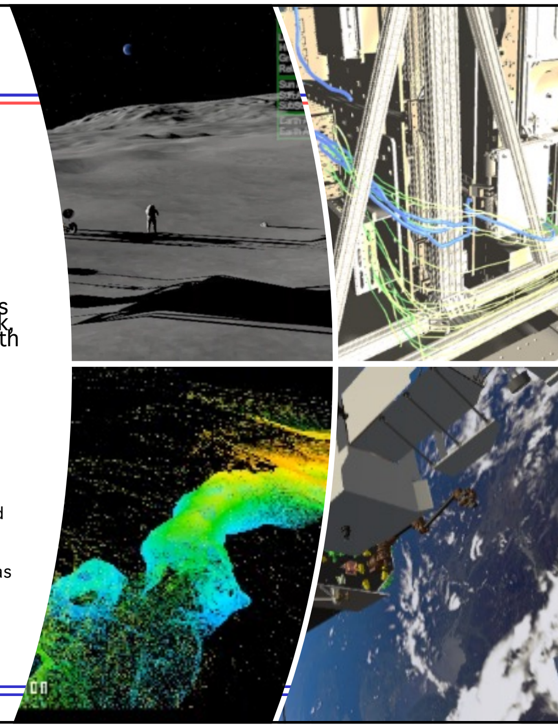
Background / Objectives



- Traditionally, scientists view and analyze the result of calculated or measured observables with static 1-D, 2-D or 3-D plots.
  - Difficult to identify, track and understand the evolution of key features due to poor viewing angles and the nature of flat computer screens.
  - Numerical models, such as the NASA GEOS climate model, are almost exclusively formulated and analyzed on Eulerian grids with points fixed in space and time.
- However, atmospheric phenomena such as convective clouds, hurricanes and wildfire smoke plumes move with the 3-D flow field, and it is often difficult and unnatural to understand these phenomena in an Eulerian reference frame as opposed to the Lagrangian reference frame in which nature operates.



Background / Objectives



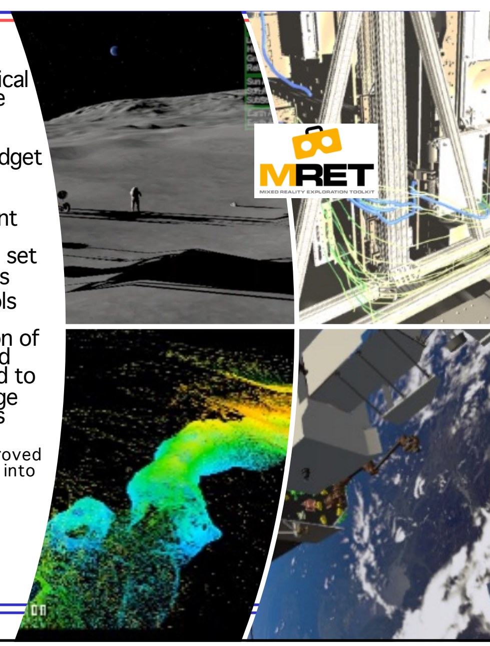
We propose to develop a scientific exploration and analysis mixed augmented and virtual reality tool with integrated Lagrangian Dynamics (LD) to help scientists identify, track, and understand the evolution of Earth Science phenomena in the NASA GEOS model. VALIXR will:

- Enhance GEOS to calculate Lagrangian trajectories of Earth Science phenomena and output budget terms (e.g., momentum) and parcel attributes (e.g., temperature) that describe their dynamics
- Enhance the NASA open source eXtended Reality (XR, i.e., AR and VR) tool, the Mixed Reality Exploration Toolkit (MRET) developed by the PI, to visualize and animate GEOS fields as well as initialize and track LD features (i.e., parcel trajectories)

ESTO Earth Science Technology Office

Technology Approach

- Integrate Giga-parcel Trajectories (GigaTraj) software into GEOS numerical weather model to calculate Lagrangian trajectories of specific Earth Science phenomena and output budget terms that describe their dynamics
- Create an open-source point cloud component for visualizing and animating a set of point clouds netCDF files
- Enhance MRET with LD tools for the initialization, finalization and visualization of trajectories with associated dynamic budget terms, and to output a parameter package for use in GEOS model runs
  - Working closely with two universities to infuse improved human factors paradigms into the visualizations and interactions in XR



6

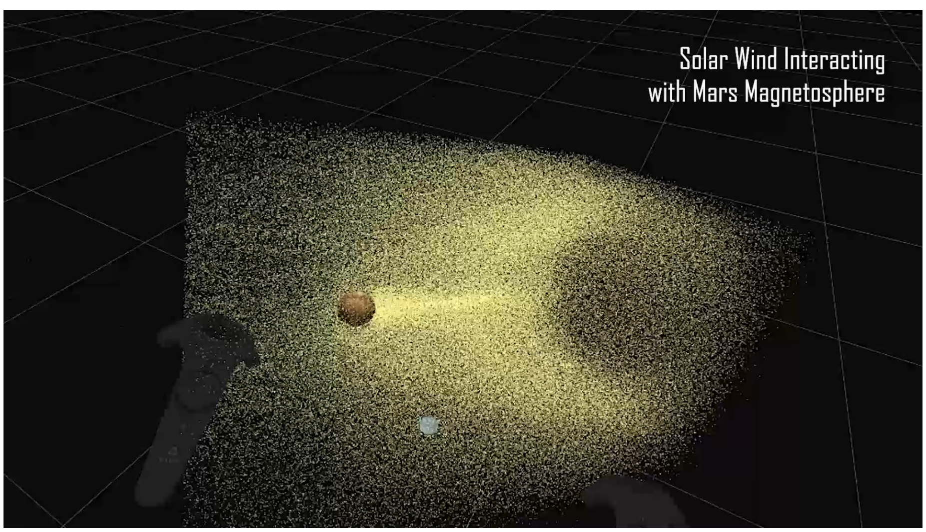
Objectives

- VALIXR will provide Earth scientists:
  - Enhanced scientific discovery of key phenomena in the Earth system through the combination of advanced visualization and quantitative LD with NASA models and data
  - An immersive, interactive, and animated visualization of GEOS fields and particle trajectories to allow scientists to intuitively initialize LD for subsequent GEOS model runs
  - **Intuitive initialization, manipulation and interaction with GEOS data and trajectory paths through the use of XR**

To provide the tools to conduct "what-if" investigations that can result in actionable predictions

ESTO Earth Science Technology Office

7



ESTO Earth Science Technology Office

8

Presentation Contents

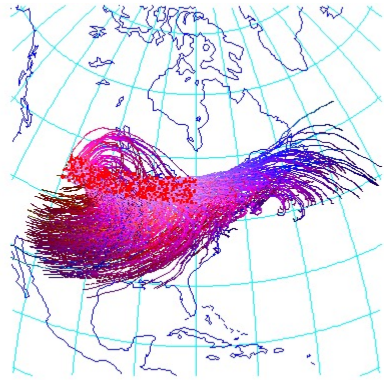
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

5 Major Components of VALIXR Work

- Integrate Giga-particle Trajectories (GTraj) software into GEOS to calculate Lagrangian trajectories of specific Earth Science phenomena and output budget terms that describe their dynamics
- Create VALIXR XR Tool:
  - Enhance the Mixed Reality Exploration Toolkit (MRET) with LD tools for the initialization, finalization and visualization of trajectories with associated dynamic budget terms
  - Create an open-source Point Cloud visualization component for MRET - Useful across NASA domains
  - Make it intuitive to understand and use - Human Factors
- Apply VALIXR to Earth Science phenomena (2 Use Cases: smoke plumes and overshooting convection among others)

Integration of GigaTraj with GEOS

- GigaTraj provides the science-rich Lagrangian data that is visualized within VALIXR: large numbers of particles, each bearing multiple physical quantities to be examined, over simulation times of hours to weeks.
- GigaTraj successfully integrated into GEOS:
  - Has access to data at each model time step (less time interpolation needed)
  - Has access to all internal GEOS model variables
  - Currently performing detailed validation against standalone GigaTraj



5-day forward trajectories from GigaTraj, initialized at 2021-07-15T12. Altitude is color-coded by hue, and age of the trajectory is color-coded by lightness. Initialization particle locations (randomly distributed within a rectangle of interest) are shown as bright red points

ESTO Earth Science Technology Office

Enhance Point-Cloud Rendering through Neural Network Techniques

Holes and Gaps

Aliasing Artifacts

- Deferred rendering with 2D neural networks can enhance high-quality and real-time rendering of massive point clouds (~1B points)
  - Mitigate image artifacts in point-cloud rendering through neural network techniques
- Leverage same concept for efficient, high-quality visualization of point clouds in VALIXR

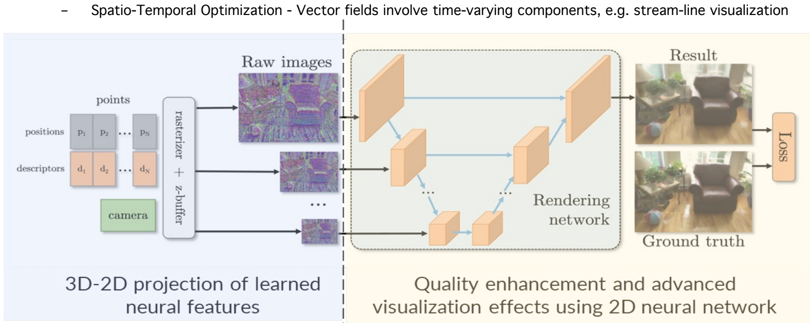
ESTO Earth Science Technology Office

Schütz, Markus, Bernhard Kerbl, and Michael Wimmer. "Rendering point clouds with compute shaders and vertex order optimization." *Computer Graphics Forum*. Vol. 40. No. 4. 2021.

Source: Aliev, Kara-Ali, et al. "Neural point-based graphics." Computer Vision, ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16. Springer International Publishing, 2020.

Point Clouds Plugin

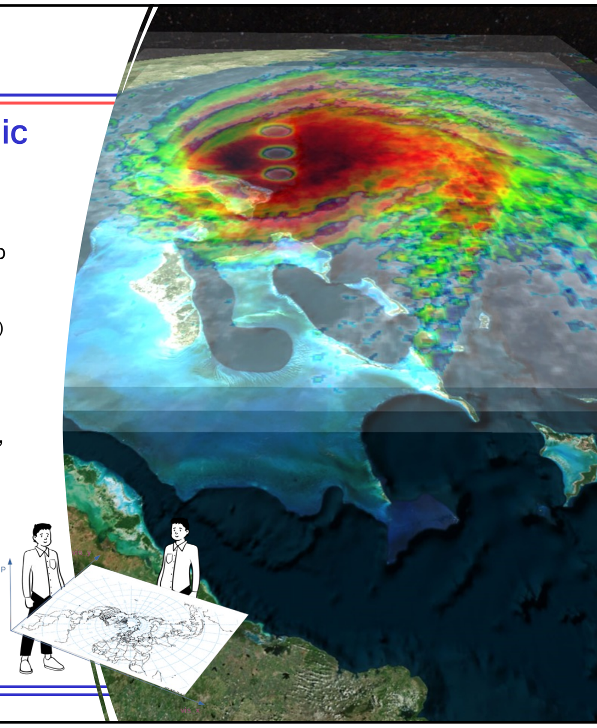
- Base Point Cloud Plugin Implementation
  - Integrated open source point cloud implementation (Potree) able to handle hundreds of millions of points (static)
    - Foundation for neural network work by UMD
  - Added netCDF support
  - Need to add temporal support for lower resolution point clouds appropriate to VALIXR use cases
- [OPTIONAL] Neural Net component for Point Cloud Rendering
  - Enhance 3D point-cloud projection with deferred neural rendering in 2D
  - Increase rendering quality and enable advanced visualization effects efficiently using 2D neural network (hole filling, antialiasing, increase photorealism, stylization, superresolution, temporal effects, etc.)
  - Reduce point-cloud rendering cost - Neural-network based superresolution (useful for “filling” in point cloud results for lower resolutions GEOS runs)
  - Stylization - Learn suitable neural embeddings to enable different visualization paradigms and support human factors
  - Spatio-Temporal Optimization - Vector fields involve time-varying components, e.g. stream-line visualization



Source: Aliev, Kara-Ali, et al. "Neural point-based graphics." Computer Vision, ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16. Springer International Publishing, 2020.

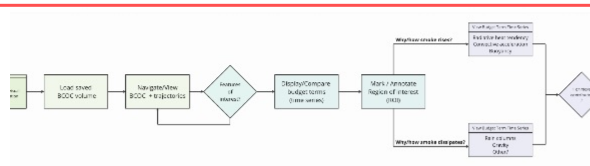
VALIXR "Holographic Table" Viz Alpha

- On-demand tiling
- Geographical/Political Maps (dynamically loaded from Web Mapping Service or REST)
  - Supports multiple levels of detail (e.g. Bing Maps ~78km/pixel to ~0.3m/pixel)
- Elevation Data (dynamically loaded from Web Coverage Service or REST)
- Image overlay layers on base mesh for 2D data (PNG, JPEG, GeoTiff, NetCDF)
- Layer template representing 3D data (e.g. Tropopause) with a procedural mesh
- Projection support (Vertical Perspective, Equirectangular)
- Implemented library of common image processing functions that can be applied to layers

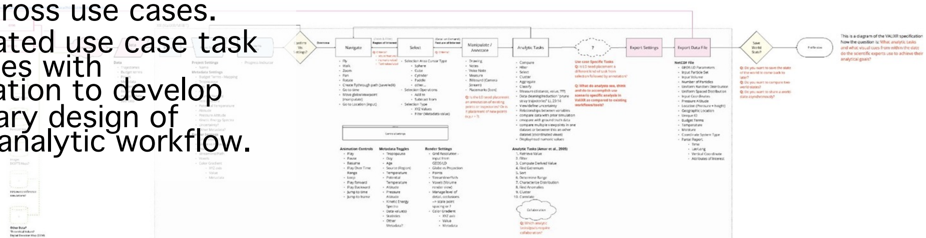


Human Factors

- Conducted Cognitive Task Analysis (CTA) with science expert Co-I's to inform human-centered design of immersive analytic workflow for VALIXR.
- Developed task flow diagrams for application use cases.
- Generated task lists for each application area use case and identified shared tasks across use cases.
- Integrated use case task sequences with specification to develop preliminary design of VALIXR analytic workflow.



Caption: CTA captured four dimensions for each use case: task sequence, decision points in the task flow along with the decision making criteria utilized, data types, visualizations types in use. Standard flow diagram symbols are utilized to represent the task sequence. Task list generated based on task sequence.



Presentation Contents

Overview (Plan)

- VALIXR B1 (Summer 2023) (TRL 4)
  - Delayed because of base point clouds implementation
  - First official version for scientists to try and give feedback
    - Basis for most Human Factors work
  - Integration of complete pipeline (TRL 4, Key, functionally critical, software components are integrated, and functionally validated, to establish interoperability)
- VALIXR B2 (Winter 2023-24) (TRL 5)
  - Use Cases (initial science use cases and human factors tests to inform new/modified features and GUI needed)
  - Human Factors
  - Feature Iteration
  - GEOS/GigaTraj Optimizations
- VALIXR Final
  - Full Use Cases
  - Feature Iteration
  - Optimizations

GEOS/GigaTraj Integration
On schedule

- Integrated GigaTraj with GEOS is complete
  - GEOS has run "successfully" with GEOS+gigatraj with MERRA2 resolution. Output has been provided to Leslie Lait for validation
  - Performing validation on the integration
- Performing cleanup and refactoring
- Bulk of optimizations will be performed in the second year. These include running gigatraj on the native cubed-sphere grid rather than on interpolated lat-lon meteorological data.

ESTO Earth Science Technology Office

24

Point Cloud Plugin
Behind schedule

- Researched scientific publications and available implementations to choose the most appropriate point cloud visualization approach for the foundation of neural network (NN)
- Base Point Cloud Plugin Implementation
  - Integrated open source point cloud implementation (Potree) able to handle hundreds of millions of points (static)
    - Foundation for neural network work by UMD
  - Added netCDF support
  - Limitations make it a bad long term fit
    - Preprocessing is needed to put point clouds into Octree format, which is unsuitable for many point cloud files for animation
    - No animation at this point
- Plans:
  - Add temporal support
  - Point Cloud Shader Implementation
    - Point cloud optimizations are done on the fly
    - Suitable for animation
    - Need to port C++ open source to C#/Unity

ESTO Earth Science Technology Office

25

VALIXR (MRET)
Behind schedule

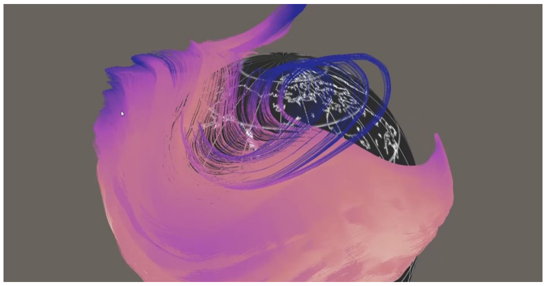
- Integrated Customized (import netCDF files to create pre-optimized and static point clouds) Potree implementation into MRET
  - Adapted Potree implementation
  - Defined solution-agnostic interface to point cloud implementations
  - Created MRET Point Cloud Manager that facilitates MRET user interactions to various (existing and future) Point Cloud components that must work in MRET, starting with the aforementioned customized Potree implementation. This System Manager will function as an abstraction between MRET and different Point Cloud implementations.
- Created GEOS Exporter for selected point clouds, which is needed for subsequent GEOS runs
- Created Volume Selection GUI (for selecting points), including sphere, cube, and cylinder
- Developed MRET "Holographic Table" alpha
- Plans:
  - Integrate all the parts together
  - Complete development of VALIXR Holographic Table
  - Add Isosurface visualization (new requirement)
  - Improve point clouds plugin
  - Iterate based on scientists' feedback and human factors

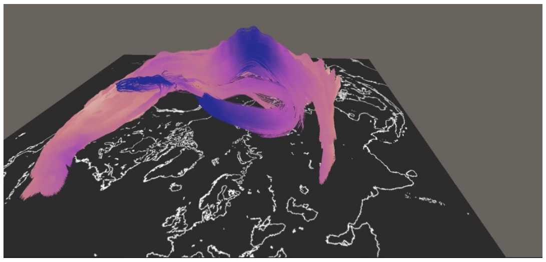
ESTO Earth Science Technology Office

26

Human Factors
On Schedule

- Human Factors is back on schedule for deliverables at end of May 2023.
- Heuristic evaluations depend on availability of trajectory and point cloud visualizations in VR and the user interface available for scientists to work with
- Conducted scientist interviews with Leslie Lait and Steve Guimond for task analysis and data abstractions
- Developed recommendations for visualization idioms
- Developed task flow diagrams for both science use cases and ValiXR





Vis idiom prelim sketches in Unity with synthetic trajectory data
Vertical: Pressure coordinates
Color encoding: warm-cool gradient: Temperature
Data: Synthetic test trajectory data from L. Lait

ESTO Earth Science Technology Office

27

Issues

- VALIXR Point Clouds Plugin Base Implementation –
  - Initial miscommunication on who had responsibility for completing the initial base point cloud implementation in the proposal, leaving the requirement uncovered
  - Initial Point Clouds Implementation has been more difficult than anticipated. Attempts to translate the compute-shader-based approach (from GLSL to Unity HLSL) were unsuccessful because Unity shader language lacks support for key features (64-bit uint, etc.). We are writing the compute-shader-based implementation from scratch starting this summer, and plan to use Unity's Custom Render Pipeline to overcome the lack of features in HLSL that are crucial for optimizing arbitrarily large point clouds during runtime.
  - Impacted prototyping, infrastructure, and human factors with delay of VALIXR Alpha 1
  - Mitigation
    - Provided additional funding to UMBC to research and create base point cloud implementation. No impact to funding because of NASA VR Team underrunning.
    - Pursued 3 different solutions
    - Separated VALIXR Alphas into 2 deliverables that are decoupled until point clouds are ready
    - Hired XR Pathways Intern in June who has Unity shader experience
- Significant delays in the delivery of grants from NSSC to the Universities, especially UMBC which took until March (7 months)
- Underrunning on NASA VR Team
  - NASA AR/VR team was understaffed and underrunning through first 3 quarters after Dylan Baker left the NASA AR/VR team before the project start date. We were unable to get approval from the center for an external hiring point to replace him. Since January, we have added 2 XR Pathways interns, 2 Civil Servants (at 0.5 FTE each), and plan to have 4 interns this Summer.
- Human Factors team was understaffed due to challenges in student recruitment in first 6 months.

ESTO Earth Science Technology Office

28

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

29

Infusion Potential

Integrating GigaTraj with the GEOS model will enable a more natural, feature-specific analysis of Earth science phenomena compared to the current Eulerian (fixed grid points in space/time) nature of the GEOS model and its associated output.

Leveraging the NASA open source MRET tool and integrating it with a generalized open-source point cloud system has huge applicability to any Earth Science domain.

- Lagrangian Dynamics and VALIXR visualizations has wide applicability in Earth Science, as well as even in Planetary Science and Astrophysics with appropriate back-end changes
  - Advances scientific understanding or diagnosis of the internal calculations in GEOS
  - Integrating the trajectory model with the GEOS system itself will provide access to such things as budget terms and dynamical feedback mechanisms that would otherwise be inaccessible
  - Unlike some trajectory models, the GigaTraj model can trace parcels kinematically (i.e., using vertical wind fields), isentropically (along surfaces of potential temperature), or diabatically (approximately along isentropic surfaces, while using diabatic heating fields to move the particles vertically), as conditions warrant

We expect that users of the GEOS modeling system, including members of this proposal, will submit new proposals to NASA and possibly other agencies that will extend the life of the LD tool beyond that of AIST.

ESTO Earth Science Technology Office

30

Infusion Potential

VALIXR MRET will provide an open source XR foundation for integration with Earth Science Digital Twin architectures and other weather (or really any point cloud generating) numerical models, and will be especially suited for the coming generation of fast turn-around, ML-enhanced numerical models.

An open-source XR software for animating and visualizing large point clouds has applicability for further research projects and applications even outside of NASA domains.

- By leveraging MRET as a foundation for VALIXR, VALIXR has applicability in domains in both science and engineering. In addition, this project can take advantage of MRET funding received from missions for engineering domain tasks (i.e., outside the scope of this proposal) to help advance the overall maturity of the AIST application. **This will improve the chance of transitioning this work from its exit TRL 6 to higher TRLs with little to no additional funding.**
- We expect that students at UMD will continue developing the framework to continuously improve performance, add functionality, and integrate it into additional applications.
- We will continue to research effective point-cloud visualization techniques targeted at specific applications, and devising novel neural network-based data representation and rendering methods that advance the state of the art in terms of the point cloud sizes that can be visualized efficiently.

ESTO Earth Science Technology Office

31

Infusion Potential (Letters of Endorsement)

- Dr. Rei Ueyama, Research Physical Scientist, Atmospheric Sciences Branch, Earth Science Division, NASA ARC
  - As a frequent user of trajectory models for both atmospheric science research and airborne mission forecasting and flight planning support, I can attest to the value that VALIXR can bring to scientists by improving the visualization (and thereby the interpretation of) air mass transport and evolution... Trajectory models are frequently used to identify the source region and history (e.g., convective influence) of sampled air parcels, which are critical for placing the *in situ* measurements in a larger atmospheric context... would be extremely valuable for visualizing and identifying the time and location of the densest part of the outflow plumes that we can target with the NASA high altitude ER-2 research aircraft.
- Dr. Nickolay Krotkov, Physical Research Scientist, Earth Sciences Division, NASA GSFC
  - A novel concept of VALIXR is that in addition to calculating point trajectories that can be continually initialized in GEOS, model equation terms such as SO2 sources/sinks and volcanic ash radiative tendencies can be output along the trajectories. The VALIXR tool could be also used for numerous air-quality applications, such as connecting pollution source emissions and dry/wet deposition regions.
- Dr. Ryan M Stauffer, Research Physical Scientist, Atmospheric Chemistry and Dynamics Laboratory, NASA GSFC
  - VALIXR's intuitive visualizations and trajectory-initialization scheme will allow us to quantify the relationships among ozone, water vapor, convection, and pollution like never before in our 24+ year record.
- Dr. Brent Garry, Planetary Geology, Geophysics, and Geochemistry Lab (Code 698), NASA GSFC
  - VALIXR has cross-over application to planetary science for visualization of 3D planetary data sets and models. These include the animation of lava flowing across lunar and martian landscapes, models of plumes on planetary bodies (i.e., explosive ash plumes from volcanoes on Mars, water-ice plumes on Europa, and sulfur plumes on Io), and solar wind interaction with the lunar surface during robotic and crewed missions.

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Publications

- **Journal / Conference Papers**
  - **Grubb, T., Clune T., Lait L., Guimond S., Engel D., Zwicker M., Eastman R., West R.,** "GEOS Visualization And Lagrangian dynamics Immersive eXtended Reality Tool (VALIXR) for Scientific Discovery Lightning Talk," Earth Science Digital Twins (ESDT) Workshop, Washington D.C., Oct 27, 2022
  - **Grubb, T., Clune T., Lait L., Guimond S., Engel D., Zwicker M., Eastman R., West R.,** "Using XR For Improving Scientific Discovery With Numerical Weather Models**,"** International Geoscience and Remote Sensing Symposium (IGARSS) Earth System Digital Twin session Paper, Pasadena, CA, July 16-21, 2023

ESTO Earth Science Technology Office

Acronyms

List of Acronyms

AAS	American Astronomical Society
ACM	Association for Computing Machinery
ADAPT	Advanced Data Analytics Platform
AIAA	American Institute of Aeronautics and Astronautics
AIST	Advanced Information Systems Technology
API	Application Programming Interface
AR	Augmented Reality
CIF	Center Innovation Fund
COTS	Commercial Off-the-shelf
DEM	Digital Elevation Map
DCOTSS	Dynamics and Chemistry Of The Summer Stratosphere
ESTO	Earth Science Technology Office
GIS	Geographic Information System
GMAO	Global Modeling and Assimilation Office
GMAT	General Mission Analysis Tool
GMSEC	Goddard Mission Services Evolution Center
GSFC	Goddard Space Flight Center
GUI	Graphical User Interface
HUD	Heads Up Display
I/O	Input/Output
IEEE	Institute of Electrical and Electronics Engineers
IRAD	Internal Research & Development
IS&T	Information Science and Technology
JPL	Jet Propulsion Laboratory
LIDAR	Light Detection and Ranging
LD	Lagrangian Dynamics
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCCS	NASA Center for Climate Simulation
NSPIRES	NASA Solicitation and Proposal Integrated Review and Evaluation System
PCD	Point Cloud Data
PY	Program Year
ROSES	Research Opportunities in Space and Earth Sciences
SOTY	Software of the Year
TRL	Technology Readiness Level
UMBC	University of Maryland Baltimore County
UMD	University of Maryland (College Park)
UMIACS	University of Maryland Institute for Advanced Computer Studies
UNT	University of North Texas
VR	Virtual Reality
XR	Extended Reality (AR/VR/MR)

ESTO Earth Science Technology Office

Reproducible Containers for Process-oriented Collaborative Analytics

Tanu Malik(PI, Organization)
Ashish Gehani (Co-I, SRI)
Kwo-Sen Kuo (Co-I, Bayesics.com)
Michael Rilee (Co-I, Bayesics.com)
Jiun-Dar Chern (Co-I, University of Maryland, College Park)

AIST-18/21-0095 Annual Technical Review
06/23/2023

Team listing: Rohan Tikmany, Aniket Modi, Moaz Reyad/DePaul University
Niklas Griessbaum, Mike Bauer, Dai Hai Ton That, Bayesics, Inc

Reproducible Containers for Process-oriented Collaborative Analytics

PI: Tanu Malik, DePaul University

Objective

Develop and provide data-efficient containerization technology for use in the precipitation features (PF)-based Earth System Digital Twin (ESDT) to conduct process-oriented model diagnostics.

Goals are to improve shareability, reusability, and reproducibility of the precipitation feature- based ESDT.

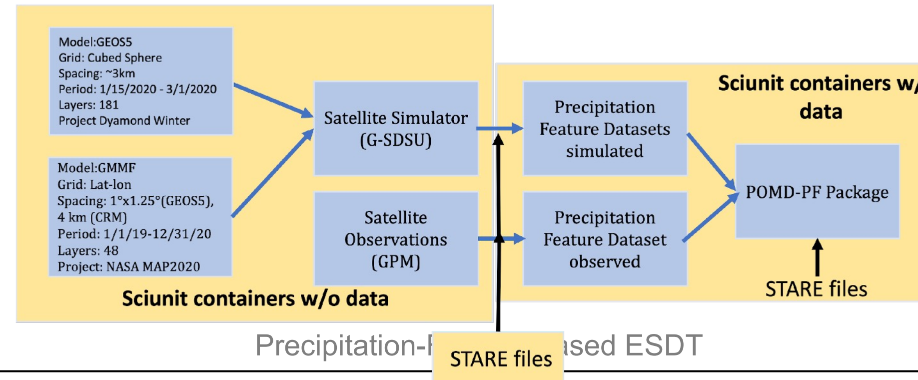
Specific **performance goals** are reducing the size of containers and enable efficient parallel analytics on the cloud with reduced size containers.

Technologies include:

- *Sciunit*, an NSF-funded containerization technology,
- *STARE*, a NASA-funded spatiotemporal library, and
- *I/O specialization*, an SRI-developed framework for reducing pre-virtualization elements.

An existing PF-based ESDT with model outputs and simulation-observation comparative analyses. (Blue boxes)

Containerizing the ESDT with Sciunit. (Yellow boxes)



Precipitation-F … ased ESDT

Approach

Modify the existing PF-based ESDT by containerizing it with

1. Sciunit containerization mechanism
2. Integrating Sciunit with (i) I/O specialization, and (ii) STARE.
3. Indexing model outputs in PF-based ESDT.

Key Milestones

- Containerize and open-source POMD-PF analysis package. 01/23
- Index model outputs in PF-based ESDT. 03/23
- Sciunit +I/O-specialized +STARE library (TRL 6) 08/23
- Demonstration use cases on PF-based ESDT 01/24
- ESDT Prototype Activity (TRL 7) 03/25

Co-Is: **A. Gehani (SRI), K. Kuo (Bayesics), M. Rilee (Bayesics), J. Chern (UMD, College Park).**

Partner: **Thomas Clune (NASA)**

TRL_{in} = 4 **$TRL_{current}$ = 4**

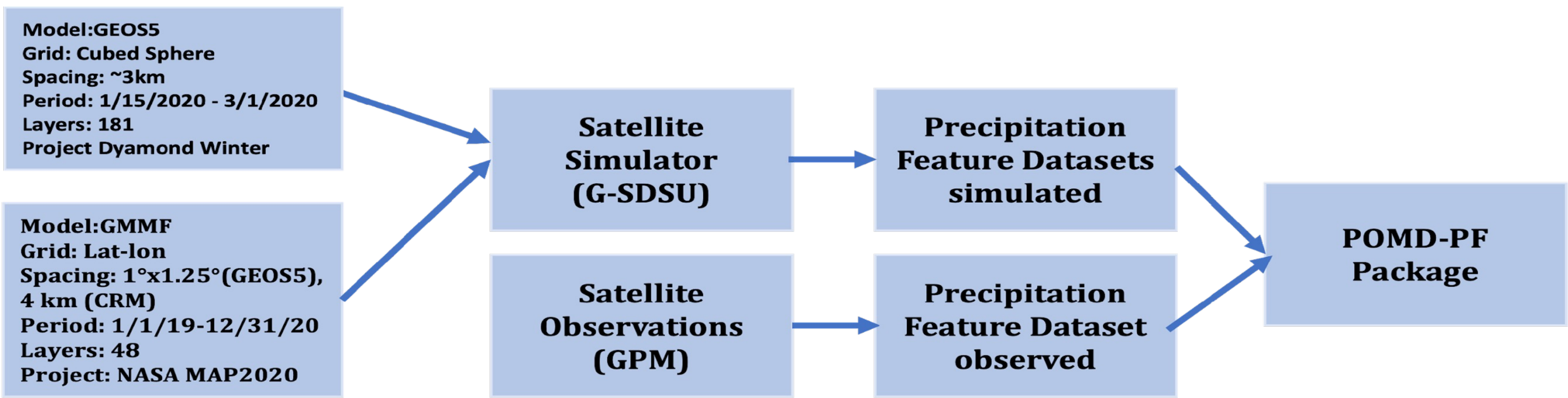
Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Background / Objectives

This project contributes to: Climate Variability and Change

Science Goal: Show how to improve Collaborative Analytics within a Precipitation feature (PF)-based Earth Systems Digital Twin



- This PF-based ESDT is used to conduct model diagnostics i.e., derive, compare and evaluate the precipitation features obtained from global storm-resolving models and observation data.
- Enables comparisons of 15 different GPM PF datasets based on various surface precipitation retrievals, sensor swaths, and brightness temperatures.

Objectives of our Project

- Identify and create collaborative analytics use-cases within a PF-based ESDT.
  - ➢ Data-intensive uses cases from **PF-based analytics package** [1]

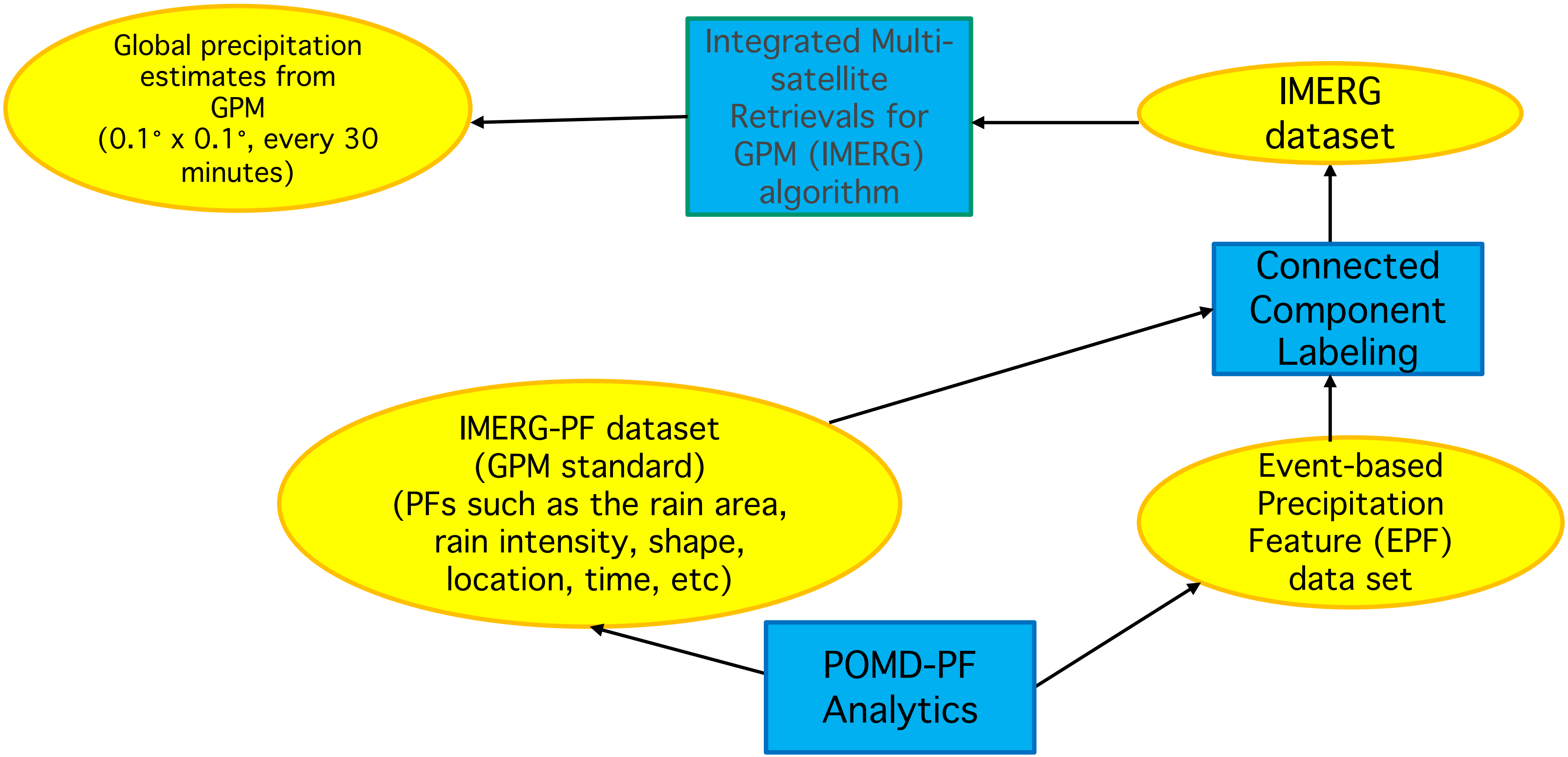
- Develop and deploy technology for sharing the use cases efficiently such that they can be reproduced exactly in different environments.
  - ➢ Use **Sciunit** [2], an automatic containerization tool for sharing and reproducing

- Determine how much of data to containerize and in which form to enable `what-if' collaborative analytics
  - ➢ Use **I/O-specialization** [3] methods to reduce the amount of data in containers
  - ➢ Use **STARE** [4] for data harmonization and homogenization.

- Develop and build an PF-based ESDT prototype based on containers for broader set of usecases.

[1] Chern, J.D., W.-K. Tao, S.E. Lang, X. Li, and T. Matsui, "Evaluating precipitation features and rainfall characteristics in a Multi-scale Modeling Framework," Journal of Advances in Modeling Earth Systems, 2020.
[2] That, D. H. T., G. Fils, Z. Yuan and T. Malik, "Sciunits: Reusable Research Objects," IEEE 13th International Conference on e-Science (e-Science), 24-27 Oct. 2017, 374-383, 2017.
[3] Niddodi, C, A. Gehani, T. Malik, J. Navas, and S. Mohan. 2020. MiDas: Containerizing Data-Intensive Applications with I/O Specialization. In Proceedings of the 3rd International Workshop on Practical Reproducible Evaluation of Computer Systems, 2020
[4] Rilee, Michael L., Kwo-Sen Kuo, James Frew, James Gallagher, Niklas Griessbaum, Kodi Neumiller, and Robert E. Wolfe. "STARE into the future of GeoData integrative analysis."Earth Science Informatics(2021): 1-18

PF-based ESDT



- IMERG-PF dataset is a valuable global survey of precipitation systems
- It addresses many scientific topics such as extreme precipitation, intense convection, moderate to heavy snow, warm rain systems, hailstorms, and the diurnal cycle, etc.

PF-based Analyses

Science Queries:

Q1. POMD-PF template that analyzes
(i) the global statistics of PFs with rain area greater than 500 km^2.
(ii) the characteristics of large storms in South America.

Q2. Find all precipitation events intersecting the San Joaquin Watershed in the month of December in 2021 and calculate summary statistics.

Q3. Compare data of different resolutions e.g., precipitation data retrieval from Level2 with satellite data e.g., kriging

Q4. Intercompare extratropical cyclones and precipitation events/feature.

Research Question:

1. How to containerize programs answering such queries?
2. How much data and in what form should be included/shared/preserved for reproducibility in such a container?

Sample Interaction

```
1. > sciunit create FIE
2. > sciunit exec ./FIE.sh ./DATA/weather_201710.Rds
     0. Download…
     1. Calculate violation matrix…
     2. Calculate heat map…
     3. Generate model data with ./DATA/weather_201710.Rds…
     4. Apply random forest model…
     5. Evaluation…
3. > sciunit list
     e1 Dec 4 12:44 ./FIE.sh ./DATA/weather_201710.Rds
4. > sciunit show
     id: e1
     sciunit: FIE
     command: ./FIE.sh ./DATA/weather_201710.Rds
     size: 306.6 MB
     started: 2017-12-04 12:44
5. > sciunit push
     ...
     Title for the new article: FIE
     new: 306.6 MB [01:05, 4.72MB/s]
6. > sciunit copy
     mSLLTj#
```

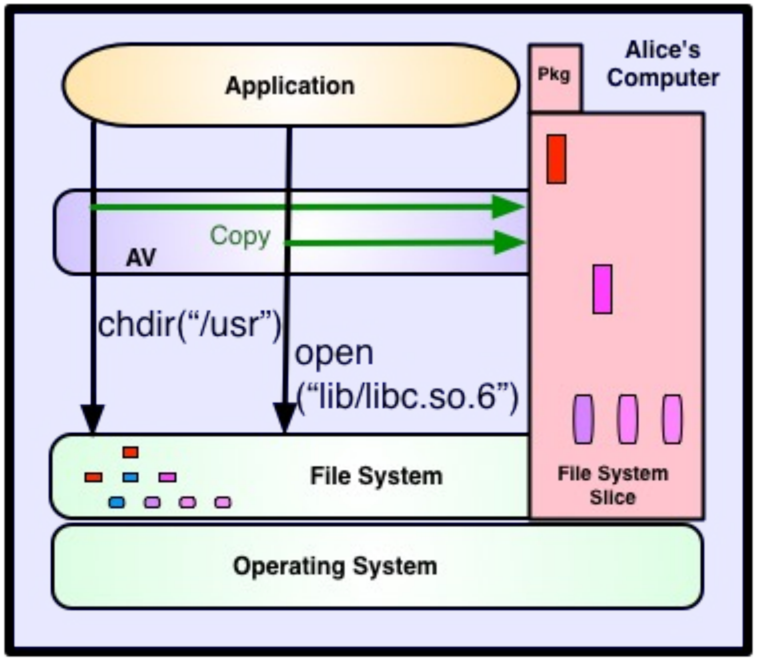
Alice's Computer

```
1. > sciunit repeat e1
     ...
     0. Download…
     1. Calculate violation matrix…
     2. Calculate heat map…
     3. Generate model data with ./DATA/weather_201710.Rds…
     4. Apply random forest model…
     5. Evaluation…
     ...
2. > sciunit repeat e1 <27050>
     ...
     3. Generate model data with ./DATA/weather_201710.Rds…
     ...
3. > sciunit given '/tmp/weather_201801.Rds' e1 %
     ...
     0. Download…
     1. Calculate violation matrix…
     2. Calculate heat map…
     3. Generate model data with /tmp/weather_201801.Rds…
     4. Apply random forest model…
     5. Evaluation…
     ...
```

Bob's Computer

D.H. Ton That, G. Fils, Z. Yuan, T. Malik. Sciunits: Reusable Research Objects. In IEEE eScience Conference (eScience), 374-383, 2017

Sciunit: Automatic Containerization of Analysis Programs

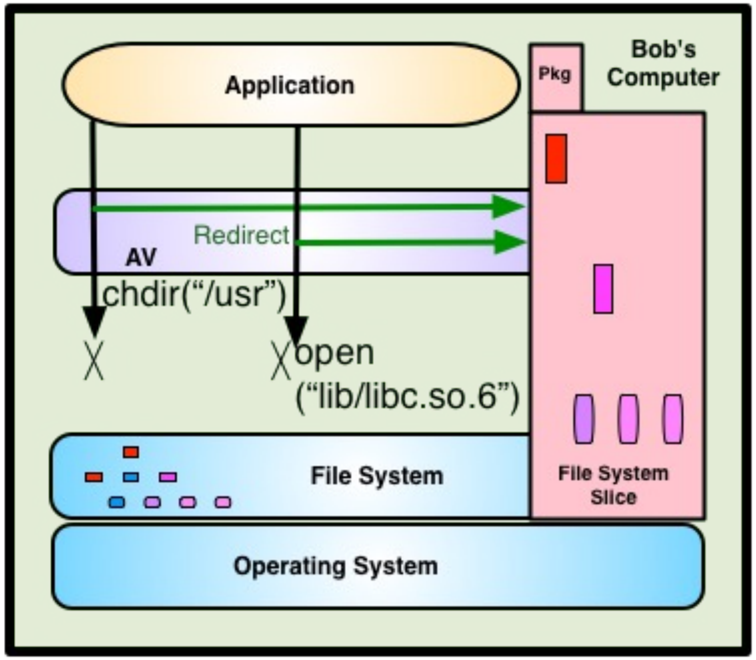


Audit Phase

- Audit ~50 system calls related to process control, file I/O, and network
  - If file is /dev/random capture return bytes as well
- At the time of interception:
  - Generate an execution trace of system call events in real-time
  - Copy files mentioned as part of system calls into a container

Q. Pham, T. Malik, B. Glavic, I.Foster. Light-weight Database Virtualization. In *IEEE International Conference on Data Engineering*, ICDE, 2015.

Target: Use *ptrace* to redirect executions into the container



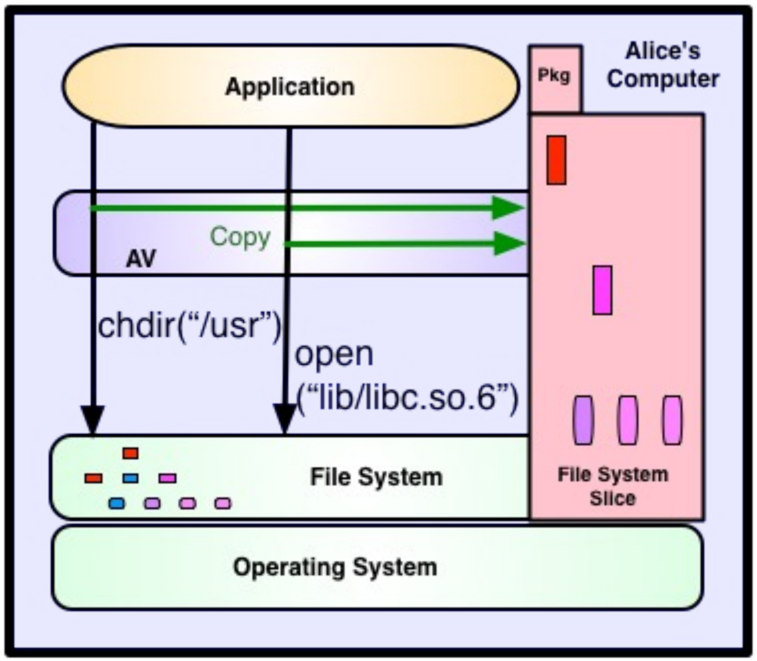
Repeat Phase

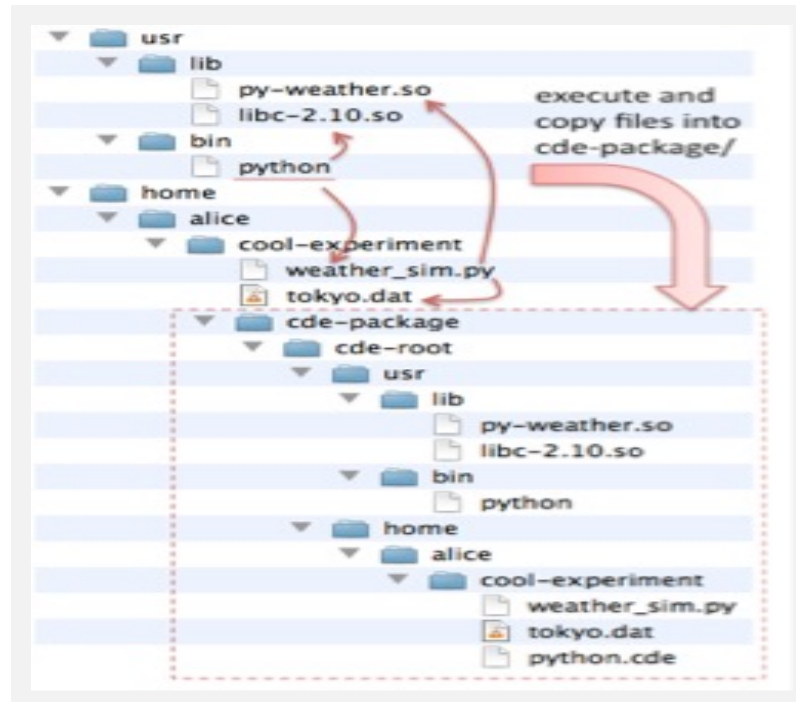
- Redirection during repetition is only for file- and network-related system events.
- Repeat execution happens within a process and file namespace.

A. Youngdahl, D.H. Ton That, T. Malik, "SciInc: A Container Runtime for Incremental Recomputation", In IEEE 15th International Conference on eScience (eScience), 291-300, 2019, doi: 10.1109/eScience. 2019.00040.

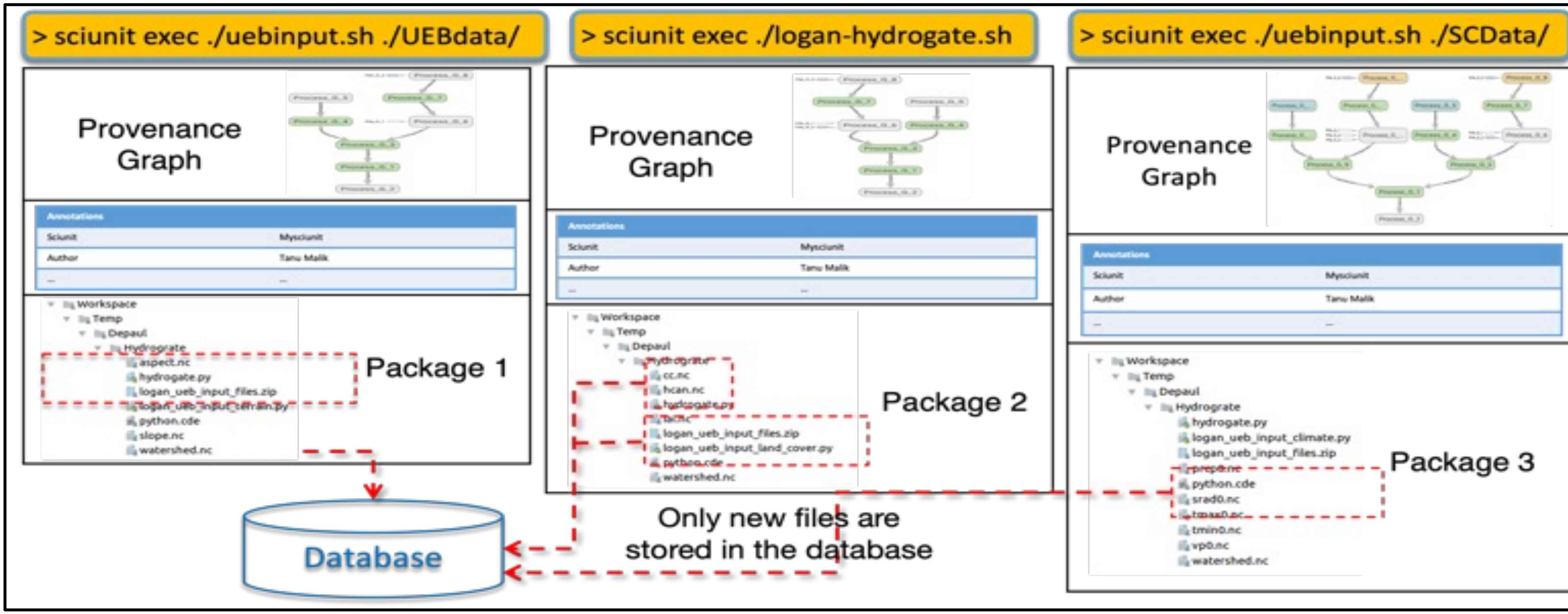
Create a *chroot*-based container

- Audit provenance during container creation time





Q. Pham, T. Malik, B. Glavic, I.Foster. Light-weight Database Virtualization. In *IEEE International Conference on Data Engineering*, ICDE, 2015.



Z. Yuan, D.H. Ton That, S. Kothari, G. Fils, T. Malik. Utilizing Provenance in Reusable Research Objects, In *Special Issue on Using Computational Provenance*, MDPI Informatics, Vol 5(1), 2018.

Alices shares *sciunits* and Bob repeats them

sciunits

container specifications
e.g., Dockerfiles

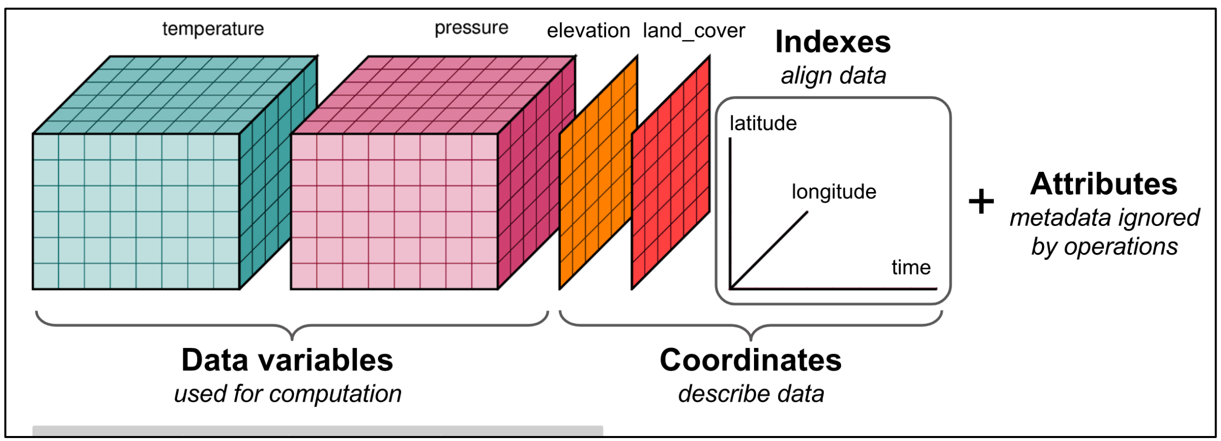
virtual environments
e.g., Conda and Venv

package manager specs
e.g., Spack files

J. Chuah, M.Deeds, T. Malik, Y. Choi, J. Goodall, "Documenting Computing Environments for Reproducible Experiments", In Parallel Computing: Technology Trends, 756-765, 2020, doi: 10.3233/APC200106.

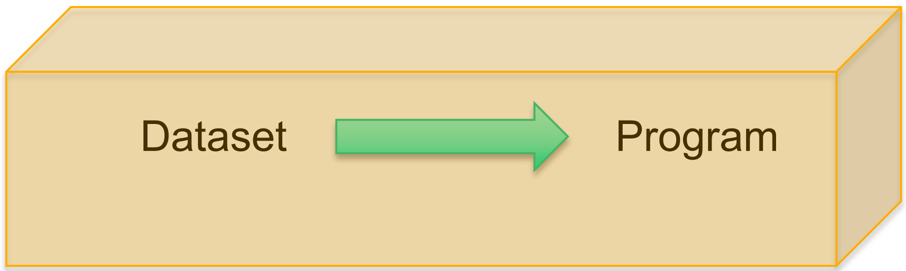
I/O Specialization

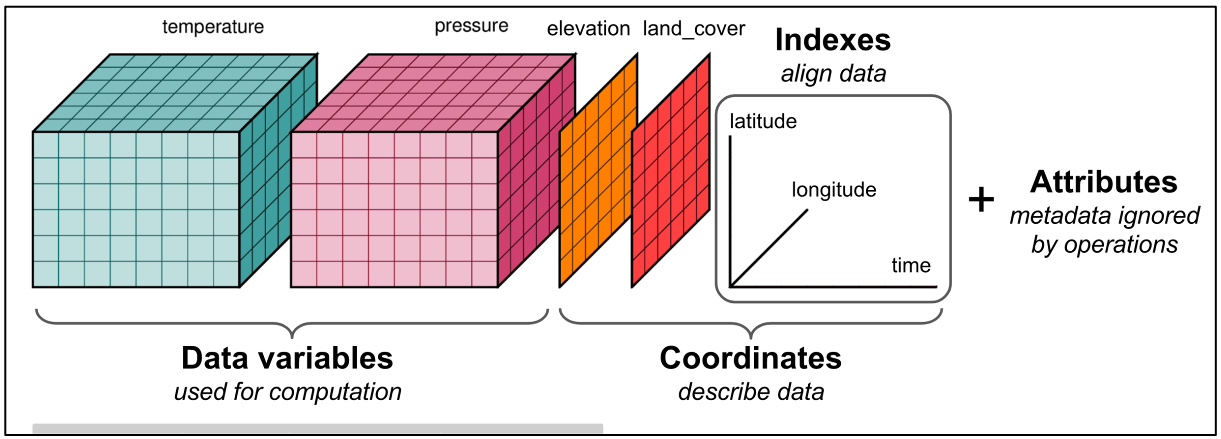
Dataset → Program



Niddodi, C, A. Gehani, T. Malik, J. Navas, and S. Mohan. 2020. MiDas: Containerizing Data-Intensive Applications with I/O Specialization. In P-RECS '20: Proceedings of the 3rd International Workshop on PracticalReproducible Evaluation of Computer Systems

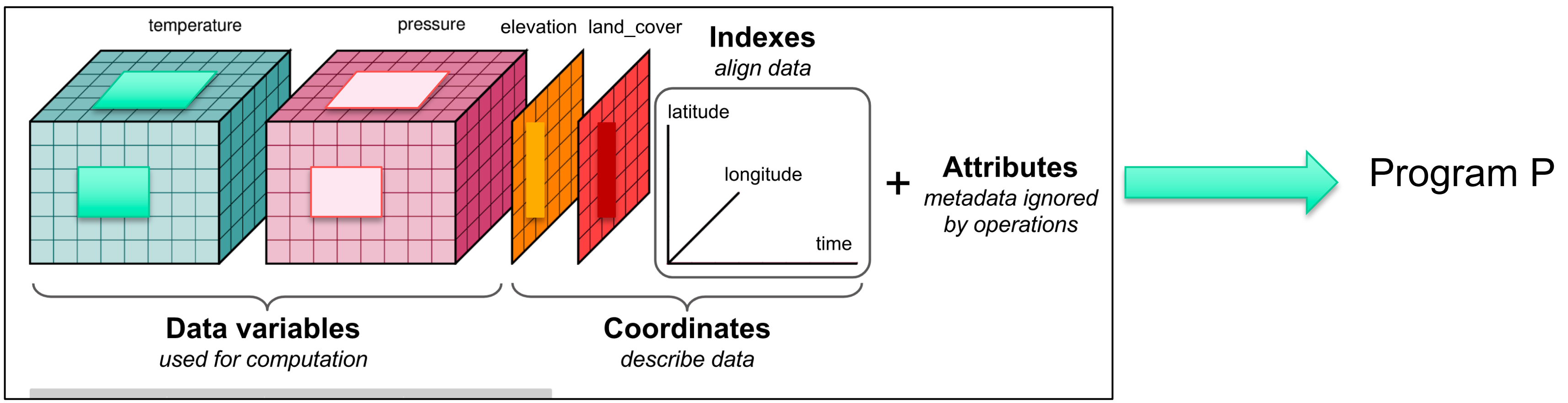
I/O Specialization





Chaitra Niddodi, Ashish Gehani, Tanu Malik, Sibin Mohan, and Michael Rilee,
IOSPReD: I/O Specialized Packaging of Reduced Datasets and Data-Intensive Applications for Efficient Reproducibility,
IEEE Access, 2023.

I/O Specialization



Research Questions:

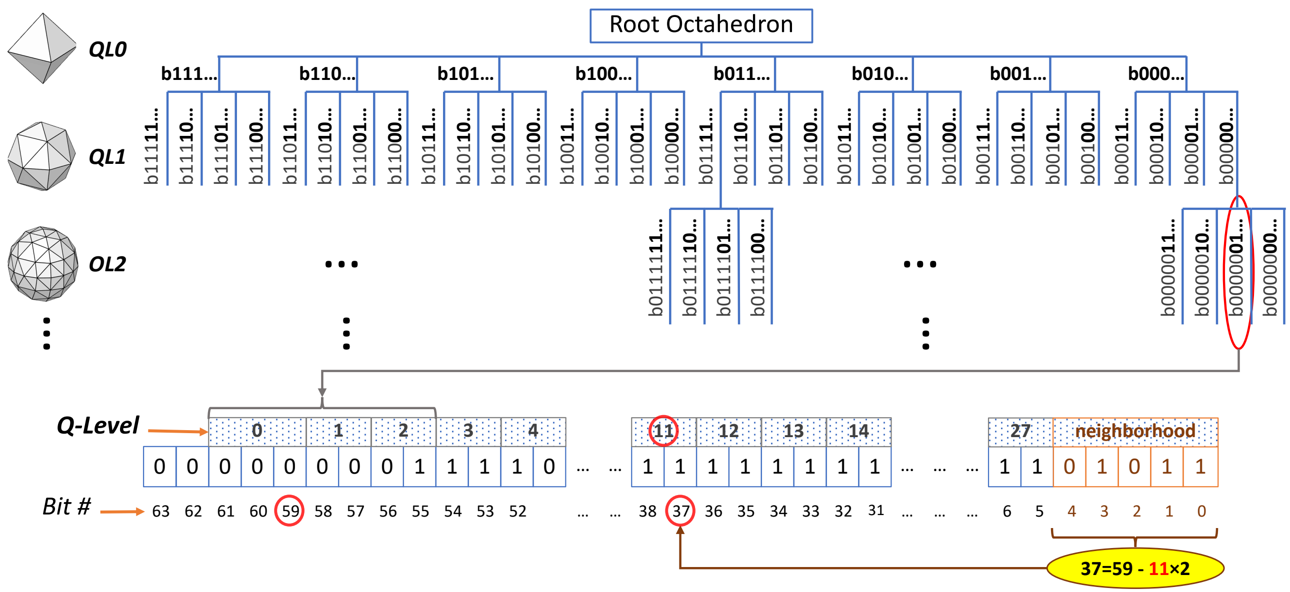
(i) Location of data files: within container or cloud-based?

(ii) Format-compliance: ability to read and use files for other purposes

(iii) Generalized dataset: total amount of data accessed by program P

STARE: SPATIOTEMPORAL ADAPTIVE RESOLUTION ENCODING

- GEO-SPATIOTEMPORAL LOCATION AND NEIGHBORHOOD
  - AKIN TO A POSTAL ADDRESS AND A ZIP CODE



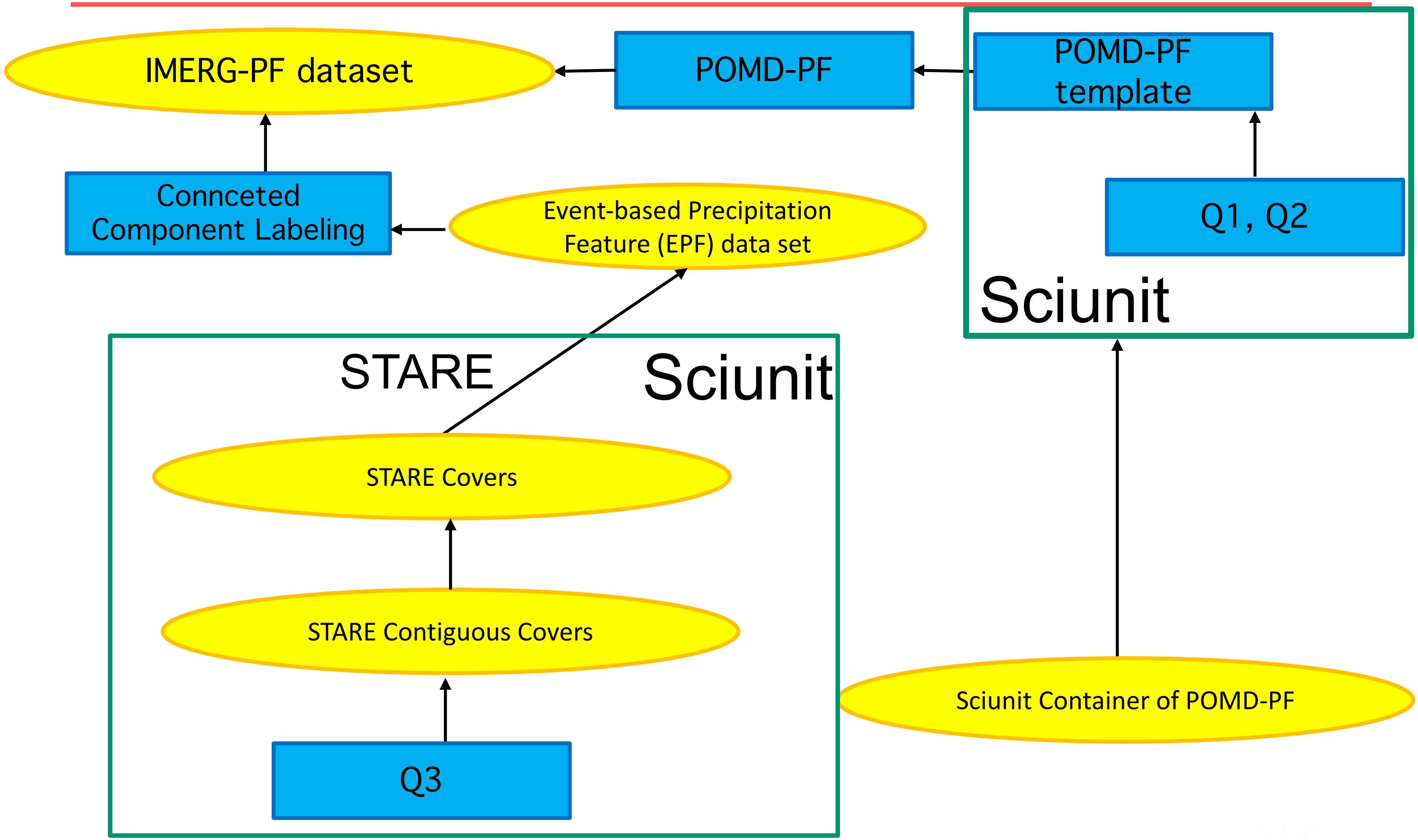
	Longitude	Latitude
Human readable	+123.4°	60°
Single-precision floating-point	0x42f6cccd	0x42700000
STARE id*	0x36ee9398f7210f34	

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Earth Science Technology Office

Primary Finding #1: POMD-PF ESDT real use cases are containerizable using Sciunit



Containerization Experience with Use-cases

- 4 uses cases ranging from statistical analysis, interpolation, and spatio-temporal comparisons of extratropical cyclones.

- No source code modification
  - Currently parallel executions using Dask are not possible.
  - We execute applications on single node but could be multi-threaded.

- Learning from the use cases:
  - Examples are increasingly developed using notebooks.
  - Examples access large amounts of data and may generate large amounts of data.
    - Docker containers do not account for size increase.
  - Container creation is easy, but provisioning is a challenge
    - Require additional information, such as instance size to deploy and repeat containers
    - Containers can audit and provide an estimate of how much time will it take to provision

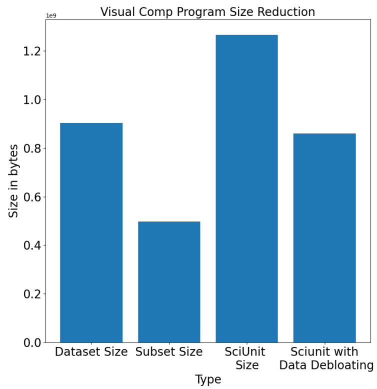
Primary Result #2: Sciunit with I/O specialization decreases size of shareable containers

Real-life applications

1. PF Features RainArea>500km^2: Visual comparison program to compare data from IMERG dataset to the data produced by GEOS-5 model for RainArea > 500 km^2
2. Krigging: A generalized interpolation method based on Gaussian regression, on the Ozone profile data from the Ozone Monitoring Instrument on the Aura Satelite
3. Global Precipitation Visualization: An ipynb notebook taken from GitHub which uses precipitation data from an IMERG dataset file to plot precipitation for different regions of the world
4. PF Features over SA: Visual comparison program to compare data from IMERG dataset to the data produced by GEOS-5 model for South America
5. Intercomparison of Precipitation Feature Events with ETC

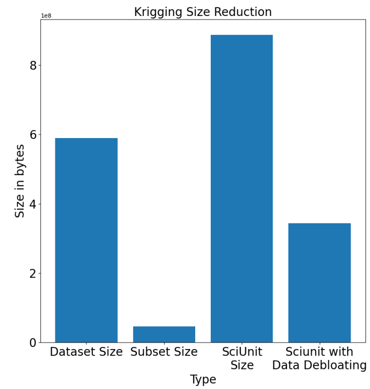
Primary Result #2: Sciunit with I/O specialization decreases size of shareable containers

PF Features RainArea>500km^2



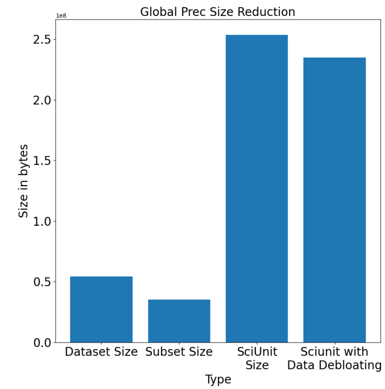
Phase	Size (MB)
Base	861.03
Subset	473.89
Sciunit	1208.31
Sciunit + Subset	821.16

Krigging



Phase	Size (MB)
Base	562.08
Subset	43.67
Sciunit	846.82
Sciunit + Data Subset	328.41

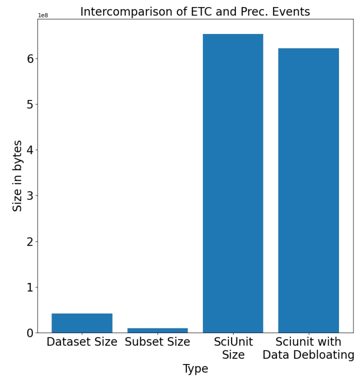
Global Precp.



Phase	Size (B)
Base	51.69
Subset	33.77
Sciunit	241.83
Sciunit + Data Subset	223.91

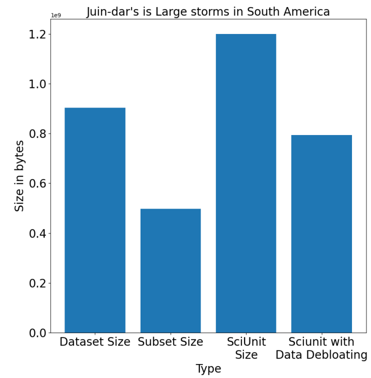
Primary Result #2: Sciunit with I/O specialization decreases size of shareable containers

InterComparison of Prec. Events with ETC



Phase	Size (MB)
Base	41.94
Subset	10.24
Sciunit	653.4
Sciunit + Data Subset	621.70

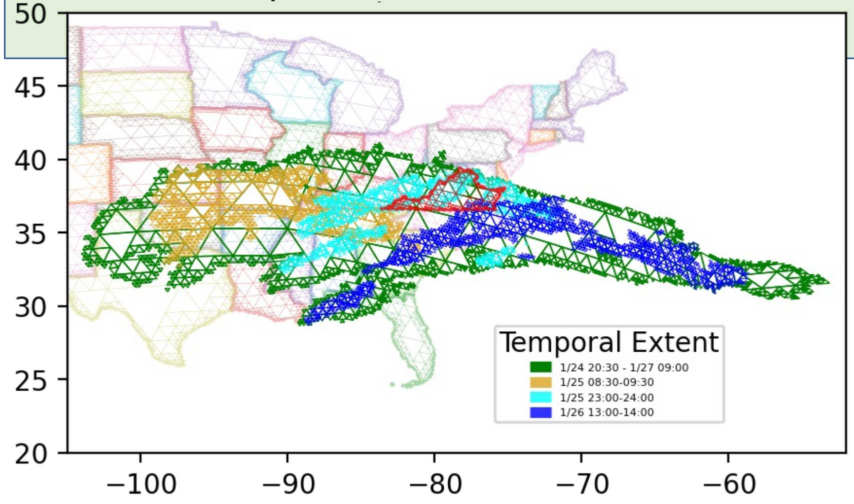
PF Features over SA



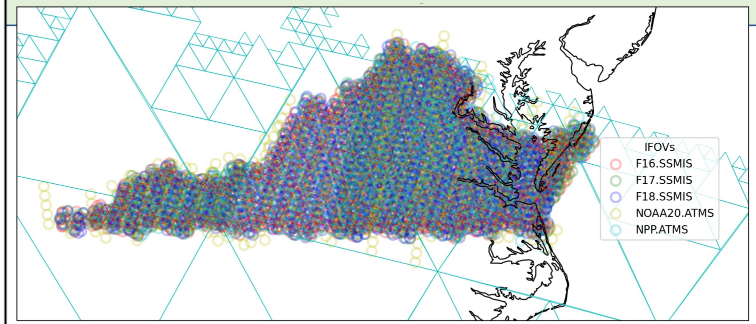
Phase	Size (MB)
Base	861.03
Subset	473.89
Sciunit	1208.31
Sciunit + Data Subset	821.16

Primary Finding #3: Enables Event Analysis which Results in Refined Precipitation Features

Time slices of the event superimposed on the total spatial extent of the event



Showing diverse XCAL datasets harmonized and subsetted for a precipitation event.



Instantaneous Fields-of-View for multiple platforms harmonized with STARE.

- Many-to-many comparison between events and features
- Subsetting of data based on STARE index using a region of interest

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Work Plan

Year 1: Requirements, Development, and Basic Containerization

~~1) Covert initial POMD-PF package from IDL to Python (*Chern, Kuo and Rilee*)~~

~~2) Develop I/O specialized libraries for NetCDf5 and HDF5.~~ Document API. (*Gehani and Malik*)

~~4) Apply Sciunit containers to POMD-PF package,~~ develop API for I/O specialized library. (*Malik*)

5) ~~Indexing observed and simulated PF databases using STARE~~ and support tooling. (*Rilee and Kuo*).

5) Deliver a data-efficient container that is good for sharing with collaborators (*Malik, Gehani*)

~~6) Produce publications and write report. (*Malik, Gehani, Chern, Kuo and Rilee*)~~

Work Plan

Year 2: Deliver Data-efficient Sciunit container for STARE-indexed PF-based ESDT

1) Integrate data-efficient containers within the POMD-PF package, perform experiments. Establish and show container reusability (*Malik, Gehani, Chern, Kuo, Rilee*)

2) Additional last mile homogenization using STARE. *(Rilee and Kuo)*

3) Integrate data harmonization layer with Sciunit to enable reproducibility. (*Malik, Kuo, Rilee, Gehani*)

4) Continued improvement of the POMD-PF package with additional diagnostics contributed by the community. (*Chern*)

5) Host Sciunit packages via cloud and S3. Create a webpage for this project which disseminates the sciunits/containers via NASA cloud, and displays display the plots from diagnostic modules for easy model-observation comparison. *(Malik, Kuo and Chern)*

6) Develop Sciunit machine (*Malik*).

7) Data partitioning for GDSU simulator. *(Rilee, Kuo and Malik, Clune)*

8) Produce publications and write report. *(Malik, Gehani, Chern, Kuo and Rilee)*

Summary of Accomplishments

- Current state and Summary of Accomplishments:
  - Several collaborative use cases in the PF-based-ESDT have been containerized.
  - An I/O-specialized Sciunit framework exists that can educate a user on the amount bloat in the container.
  - User can use configuration parameter to decide to include entire dataset or reduced dataset.
  - Sciunit container will establish repeatability guarantees in different environments with carved dataset.
  - Sciunit containers will include STARE files, provided the program also accesses them.

Future Plans

- Sciunit with I/O-specialization—leading to TRL4.
- Automatic creation of STARE with Sciunit---looking at integration opportunities.
- Multi-node scheduling of Sciunit-I/O specialized containers.
- Integration with other ESDT projects.

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Actual or Potential Infusions and Collaborations

- Infusion to existing projects:
  - Precipitation features analytical packages (POMD-PF) also applies to GMAO/GEOS-5 model output.

- Infusion to other AIST/ESDT projects
  - We plan to work more closely with our collaborator Tom Clune in the second year more regularly to determine infusion with other AIST/ESDT projects
  - We are also interested in learning more about the following AIST projects
    - AIST-21-0063, AIST-21-0082, AIST-21-0012, AIST-18-0099, AIST-QRS-22-0001

- Technology Transfer
  - All components are open-source, available via public GitHub repositories licensed via MIT Open Source License

- Transition
  - PI Malik is involved with NSF EarthCube Program and the notebook-based Sciunit container technology developed via that project is being applied to the use cases identified in this project.

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

Publications

Journal / Conference Papers

1. *Chaitra Niddodi, Ashish Gehani, Tanu Malik, Sibin Mohan, and Michael Rilee*, **IOSPReD: I/O Specialized Packaging of Reduced Datasets and Data-Intensive Applications for Efficient Reproducibility**, *IEEE Access*, 2023.
2. *Aniket Modi, Moaz Reyad, Tanu Malik, Ashish Gehani*, **Querying Container Provenance,** *In ACM Theory and Practice of Provenance, 2023.*
3. *Rohan Timany, Aniket Modi, Tanu Malik, Ashish Gehani, Deepak D'souza, Raghavan Komondoor*, **Efficient data debloating,** *In preparation.*
4. *Michael Rilee, Niklas Griessbaum, Dai Hai Ton That, Michael Bauer, Kwo-Sen Kuo,* **STARE EVENT EVOLUTION (SEE) ANALYTICS,** *IEEE/IGARSS Pasadena, California 2023.*
5. *Michael Rilee, Michael Bauer, Niklas Griessbaum, Dai Hai Ton That, Kwo-Sen Kuo,* **BEYOND THE DATACUBE,** *IEEE/IGARSS Pasadena, California 2023*

Dissertations

1. R. Tikmany, Trace-based Data Debloating, MS. Thesis, DePaul University, 2023.
2. A. Modi, Access-based Data Debloating, M.Tech Thesis, CSE, IIT, Delhi, 2023.

List of Acronyms

• 1D	One-dimensional
• CMS	Content Management System.
• CSV	Comma Separated Variables
• DEM	Digital Elevation Model
• DS	Decadal Survey
• Etc.	

A scalable probabilistic emulation and uncertainty quantification tool for Earth-system models

Matthias Katzfuss (PI, Texas A&M University)
Jonathan Hobbs (Co-I/Institutional PI, JPL)
John Reager (Co-I, JPL)

AIST-21-0082 Annual Technical Review
June 19, 2023

Team members:
Paul Wiemann, Carrie Lei, Jian Cao (TAMU)
Matthew Bonas (JPL/Notre Dame)

1

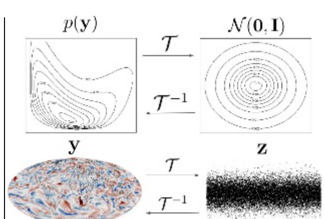
1

A scalable probabilistic emulation and uncertainty quantification tool for Earth-system models

PI: Matthias Katzfuss, Texas A&M University

Objective

- Improve mapping of uncertainties in climate predictions under what-if scenarios
- Develop statistical methodology for efficient estimation of and simulation from the probability distribution of geophysical fields
- Implement the methodology in an open-source software package with thorough documentation
- Demonstrate use of the toolbox in assessing the probability distribution of important hydrological variables, as represented in Earth system models



The transport map transforms a non-Gaussian distribution to a standard Gaussian distribution.

Equivalently, the map converts a geophysical field to standard Gaussian coefficients.

The map coefficients can be viewed as scores corresponding to a nonlinear version of principal components, and they decrease in importance and in corresponding spatial scale from left to right.

Approach

- Using Bayesian transport maps, learn the spatio-temporal and multivariate dependence structure from a small- to moderate-sized ensemble of runs from an Earth-system model
- Provide software as a fully automated package in Python, which can be easily used by other researchers
- Obtain a realistic representation of the current and future joint behavior of precipitation, snowpack, and runoff from output from land-surface models (LSMs)

Co-Is/Partners: Jonathan Hobbs, JT Reager (JPL)

Key Milestones

• Provide tech report on spatio-temporal method	02/23
• Demonstrate software on simulated example	08/23
• Apply software to LSM output	02/24
• Complete and submit manuscript on stat. methods	08/24
• Complete and submit manuscript on LSM applic.	08/24
• Make software publicly available	08/24

TRL_{in} = 3 **TRLcurrent = 3**

11/21

2

2

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

3

3

Background

Problem statement: There is much uncertainty in projections of future climate, due in part to uncertain representations of physical processes in Earth System models. The amount of data in these models can be overwhelming. Better identification and characterization of locations, times and variables of greatest uncertainty could create opportunities for higher-efficiency research and the closure of observational gaps for key processes.

- We aim to improve and simplify UQ for climate models.
- Our proposed technology can serve as a crucial component in numerous NASA applications, including studying climate projections, efficient OSSEs, UQ efforts for existing missions, and what-if investigations for potential future observing systems (i.e., mission-formulation trade-space studies).
- This project is a case study to identify impactful characteristics of the water cycle under current conditions and its response to climate change.

ESTO Earth Science Technology Office

4

4

Objectives/Goals

- G1: Develop statistical methodology for efficient estimation of and simulation from the probability distribution of geophysical fields.
  - The approach, based on Bayesian transport maps, learns the spatio-temporal and multivariate dependence structure from a small- to moderate-sized ensemble of runs from an Earth-system model.
- G2: Implement the methodology in an open-source software package with thorough documentation.
  - The software will be user-friendly, fully automated, and scalable to very large datasets.
- G3: Demonstrate the toolbox in assessing the probability distribution of important hydrological variables.
  - Based on output from CMIP6 land-surface models (LSMs), we will investigate the distributions of hydrological variables under multiple emissions scenarios to characterize their distributions under current and end-of-century conditions.

ESTO Earth Science Technology Office

5

5

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

6

6

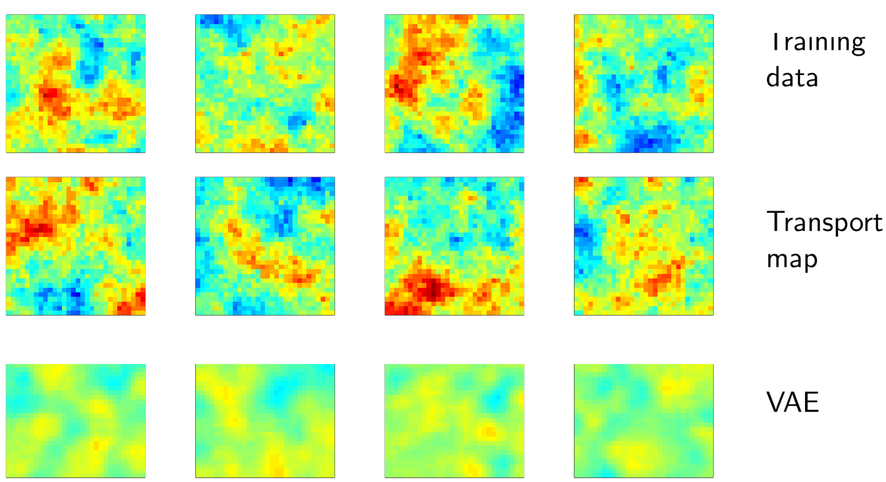
Technical and Science Advancements

- The project builds upon an approach that can learn spatial dependence structure from a small ensemble of nature runs using Bayesian transport maps.
- Extensions (in terms of methodology, software, and application) of the spatial approach:
  - to spatio-temporal fields
  - to multivariate fields considering multiple variables jointly
  - to allowing the inferred distributions to depend on fixed, known conditions or covariates (e.g., emissions scenarios, different Earth system models)

7

Emulating samples from a Gaussian distribution

Gaussian distribution with exponential cov. on grid of size $N = 30 \times 30$

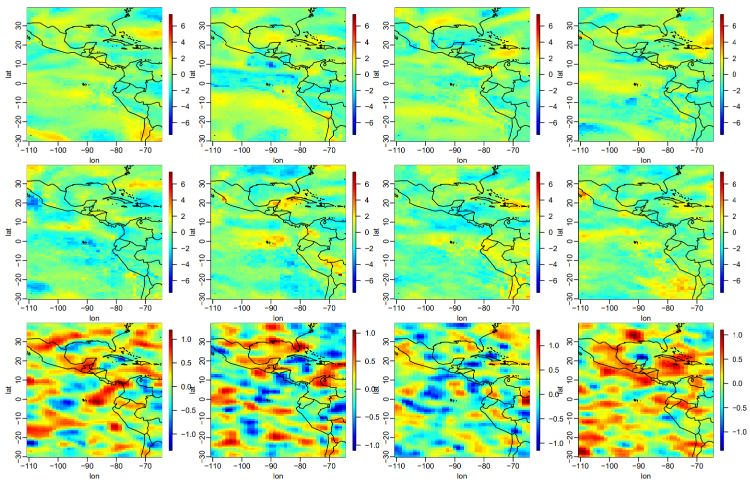


Katzfuss & Schäfer (JASA, 2023)

8

Emulating precipitation fields

Nonlinear TM for precip anomalies on subregion ($N = 37 \times 74 = 2{,}738$)



Training data

Transport map

VAE (scaled)

9

Katzfuss & Schäfer (JASA, 2023)

9

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

10

10

Current State

- Methodology development:
  - The spatial approach can learn spatial dependence structure from a small ensemble of nature runs using Bayesian transport maps.
  - The spatio-temporal extension can provide realistic-looking samples and better fit than a parametric model when trained on climate-model output.
    - First step: Scaling of space vs time learned by fitting a parametric model to the data.
    - Second step: Fit transport map in scaled space-time input space.
    - Details provided in tech report.
  - Similar initial development of extension to multivariate spatial fields.
- Software development:
  - We have an initial implementation of the spatial and spatio-temporal methodology using Python and PyTorch.
- Application:
  - The team has obtained ensemble model output from the Community Land Model (CLM) as part of the NCAR large ensemble project.
  - The ensemble output for snow water equivalent (SWE) is being analyzed as a demonstration of the spatial transport map methodology.

ESTO Earth Science Technology Office

11

11

Future Plans

- 08/23:
  - Demonstrate software on simulated examples
- 02/24:
  - Apply software to LSM output
- 08/24:
  - Complete and submit manuscript on statistical methods
  - Complete and submit manuscript on LSM application
  - Make software publicly available

ESTO Earth Science Technology Office

12

12

Presentation Contents

Actual or Potential Infusions and Collaborations

- Potential infusions/transfers
  - NASA Western Water Applications Office: potential interface for stakeholders/users of current and future hydroclimate probabilistic scenarios
  - NASA/JPL mission design technology development activities, including planetary boundary layer incubation studies
  - NASA Earth System Observatory missions, particularly the Atmosphere Observing System
- Non-NASA collaborations
  - PI collaborating with Sandia National Laboratory on transport map application for climate interventions

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Publications

- **Journal / Conference Papers**
  - Katzfuss, M., and Schäfer, F. 2023. Scalable Bayesian transport maps for high-dimensional non-Gaussian spatial fields. *Journal of the American Statistical Association*, accepted.
- **Other**
  - Tech report on spatio-temporal extension (submitted in ERS)
  - Several presentations on this work this summer at the Spatial Statistics conference in Boulder and at the Joint Statistical Meetings in Toronto

Acronyms

List of Acronyms

- CMIP6 Coupled Model Intercomparison Project, sixth phase
- JPL Jet Propulsion Laboratory
- LSM land-surface model
- OSSE observing system simulation experiment
- UQ uncertainty quantification

Kernel Flows: emulating complex models for massive data sets

Jouni Susiluoto* (PI)
Amy Braverman*
Nimrod Carmon*
Kerry Cawse-Nicholson*
Ziad Haddad*
Otto Lamminpaa*
Houman Owhadi**

* Jet Propulsion Laboratory, California Institute of Technology
** California Institute of Technology

AIST-21-0012 Annual Technical Review
June 23 2023

1

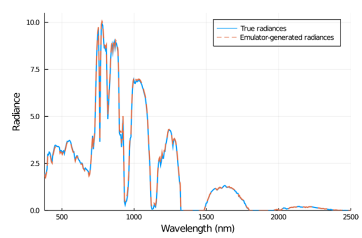
1

Kernel Flows: emulating complex models for massive data sets

PI: Jouni Susiluoto, (JPL)

Objective

- Develop an accurate, high performance uncertainty-aware emulator software and release it as open source software.
- Demonstrate the software with two applications:
  - Radiative transfer (RT) modeling, which can be used e.g. with future remote imaging spectroscopy missions such as NASA Surface Biology and Geology, part of the NASA Earth Observatory, and
  - Convective storm now-casting, which is a societally relevant research question.



Example of emulated vs. true radiance spectra in the visible and shortwave IR range from a preliminary study for the SBG mission. The spectra can be used for atmospheric correction and to retrieve surface reflectances.

Approach

- We use Gaussian process models together with cross-validation to construct the emulators.
- The specific algorithm that combines these approaches is called Kernel Flows.
- Kernel Flows is able to describe data, whose covariance changes over the domain of inputs (e.g. spatial locations).
- Cross validation ensures that the learned emulators are not overly confident and produce realistic uncertainty estimates.

Co-I's: Dr. Ziad Haddad (JPL), Dr. Kerry Cawse-Nicholson (JPL), Prof. Houman Owhadi (Caltech), Dr. Amy Braverman (JPL), Dr. Otto Lamminpaa (JPL), Dr. Nimrod Carmon (JPL)

Key Milestones

- Training data have been assembled, sampling designs and prediction models for applications are ready 05/23
- Emulators trained, software is functionally complete 09/23
- Radiative transfer emulation work integrated into retrieval framework, journal paper submitted 03/24
- Relase of final software version 05/24

TRL_{in} = 3, $TRL_{current}$ = 3

08/22 AIST-21-0012

2

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Background and Objectives

- Modeling / studying complex phenomena generally benefits from tools that either a) facilitate simulation, b) make inference easier, or c) enable straightforward lightweight data-driven modeling. **In this project we develop, implement, and deliver such a tool.** We have several application examples; these are presented in the slides below.

- The project consists of three main components:
- Deliver a documented and tested open source software package for fast, and accurate emulation for NASA science and mission purposes,
- Demonstrate the benefits of this software for radiative transfer emulation (A1) in the context of remote sensing retrieval algorithms, and
- Apply the algorithm to now-casting convective storm evolution in real-time (A2)
- Furthermore, we:
- *O2: fully utilize the large amount of . . . observations using advanced analytical tools*
- *O3: provide technology for enabling near- and long-term science and policy decisions . . . including . . . development of new models or science analysis, . . . applications to inform choices, support decisions, and guide actions for societal benefit*

Presentation Contents

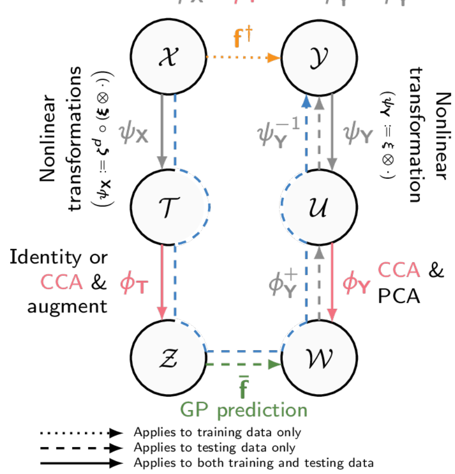
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Technical and Science Advancements 1

- We are currently 10 months into the project
- Main finding: our method works surprisingly well, and is systematically able to capture whatever structure there is in the data
- Applications: SBG radiative transfer, OCO retrievals, ice-water path retrievals, climate indexes, etc.

- Technology development: software interfaces have been defined, real-world example from OCO on next slide
- Multiple new algorithms developed and tested (some work, some don't)
- Potential new research directions for theoretical work with university partners
- Code is now open source (not uploaded for public availability yet, however)

$$\mathbf{f}^{\dagger} \approx \mathbf{f} := \psi_{\mathsf{X}} \circ \phi_{\mathsf{T}} \circ \bar{\mathbf{f}} \circ \phi_{\mathsf{Y}}^{+} \circ \psi_{\mathsf{Y}}^{-1}$$



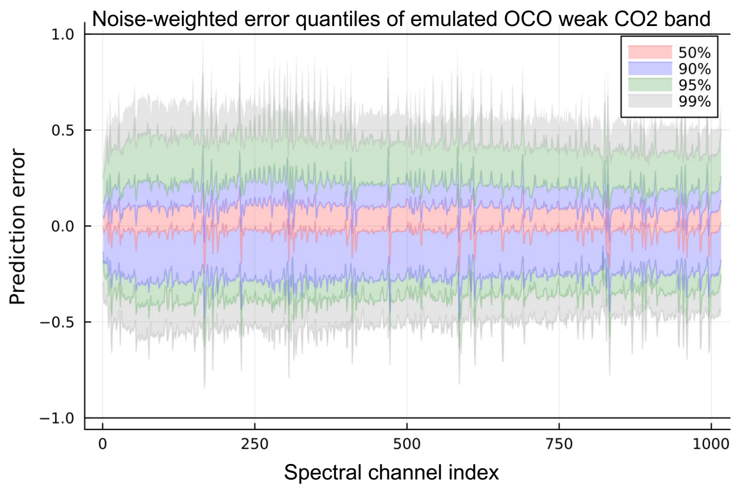
Technical and Science Advancements 2

```
using LossFunctions
using DimensionReduction
using KernelFunctions
using ParametricModel
using ParametricTraining
using ParametricPredict
using ParametricMultilevel
using OCOData

# Get data for the strong CO2 band
X_tr, Y_tr, X_te, Y_te = OCOdata(5000; "sco2")

# Get dimension reduction structs
DXs, DY = dimreduce_CCA_PCA_augmented(X_tr, Y_tr, nvecs_Y_CCA = 9, nvecs_Y_PCA=3, nvecs_X_CCA = 3, reg = 5e-3)

# Construct multivariate GP model, train and predict
MVM = MVGPModel(X_tr, Y_tr, Matern52, DXs, DY; transform_zy = false)
train!(MVM, ρ_RMSE; ngridrounds = 10, n = 16, navg = nothing, niter = 2000, ϵ = .05)
Y_te_pred1 = predict(MVM, X_te)

# Construct residual (multilevel) model, train and predict
MVM2 = MVGPModel_twolevel(MVM, X_tr, Y_tr; nvecs = 6, kernel = Matern32)
train!(MVM2, ρ_RMSE; ngridrounds = 10, n = 16, navg = nothing, niter = 2000, ϵ = .05)
Y_te_pred2 = predict(MVM2, X_te)

# Predicted values are computed by summing the levels
Y_te_pred = Y_te_pred1 + Y_te_pred2
```

Example of current API. After loading data, model setup, training, and prediction is four lines of code. Adding another GP layer adds four lines more.

7

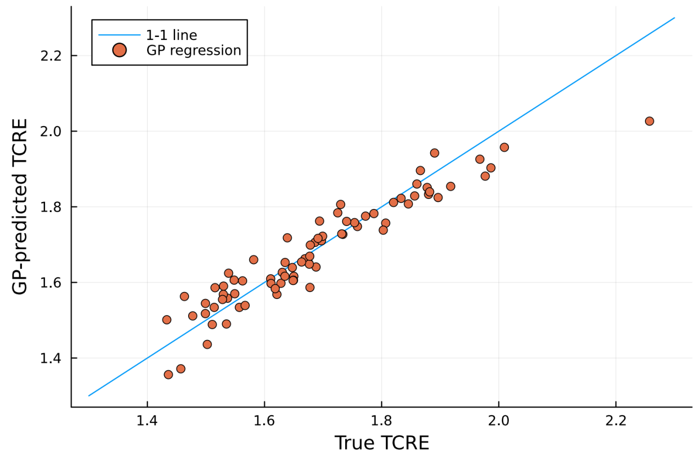
Technical and Science Advancements 3



OCO radiative transfer example: we would like to be under instrument noise level (+-1) and we are well below that. This is true also for the two other quantities (O2 and strong CO2 bands). Two-level CCA-based GP was used here.

8

Technical and Science Advancements 4

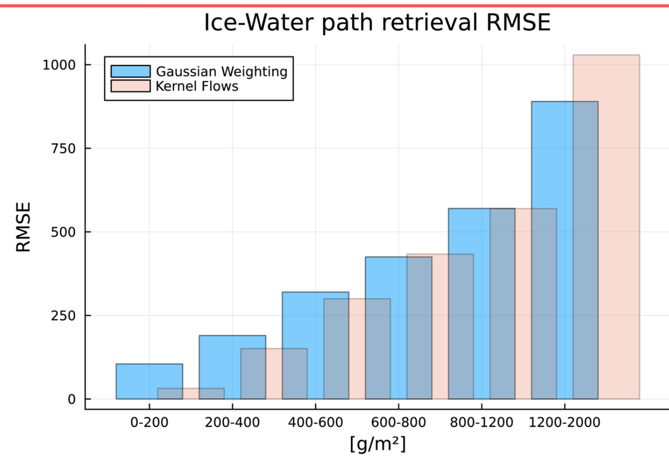


Emulation of Transient Climate Response to Cumulative Carbon Emissions (TCRE) based on 200 training data from the University of Victoria Earth System Climate Model (UVICESCM), with 77 test data points over 20 input variables and sub-optimal design. Data generation with full model took a full month on Europe's fastest supercomputer. GP results (including coding) were produced over a coffee break. (data from Antti-Ilari Partanen / Carla di Natale (FMI))

ESTO Earth Science Technology Office

9

Technical and Science Advancements 5



Application A2: We are able to improve over current Gaussian weighting-based retrieval method, but data is very noisy. A2 use case is difficult and therefore has been driving our technical innovation (as intended). Improvement at the low end is substantial and physically important. At the large end the estimates are, however, worse (it seems to be a zero-sum game to some extent). We are still talking about metrics to use.

ESTO Earth Science Technology Office

10

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

11

11

Summary of Accomplishments and Future Plans

Summary:
- Software is in good shape already now
- Open sourcing is done; code needs to be uploaded
- Accuracy exceeds / meets expectations
- Compute performance exceeds / meets expectations
- Imaging spectroscopy application is on track
- Storm nowcasting is slightly behind, but we are not really worried about that
- OCO has emerged as an important application, tested in retrievals with 1,000x-10,000x speed-up; development carried out in close contact with mission

Plan:
- Rework software into a Julia package format
- Perform imaging spectroscopy retrievals with the software
  - We have built a EMIT-type L2 retrieval framework that we'll use to demonstrate this.
  - We would like to add more complex forward model and hence showcase that our approach is really flexible
- Documentation (paper in prep)
- Continue core application work in close contact with SBG / OCO / INCUS

Risks:
- Balancing algorithmic innovation with stabilization
- A2 work has been slightly slower than we would have liked

Solution is to keep an eye on these aspects for now, dedicate resources to A2, and to keep stabilizing.

Critical issues:
- For some of our potential collaborations (Venus, mission design), graph discovery will be needed / useful
- Expanding our work to graph discovery could significantly increase our impact and allow emulation of more complex systems.

ESTO Earth Science Technology Office

12

12

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

ESTO Earth Science Technology Office

13

13

Actual or Potential Infusions and Collaborations

- Potential infusions:
- OCO as described above
- SBG as described above
- INCUS (plan is to work on this in Fall 2023)

- Ongoing / planned outside collaborations:
- Planned: Venus-related work with prof. Theo Kurten and Dr. Benjamin Frandsen (University of Helsinki)
- Ongoing: climate studies, with Dr. Antti-Ilari Partanen and MSc Carla di Natale (Finnish Meteorological Institute)
- Planned: Kernel methods for L3 CO2 products; Dr. Marko Laine and Dr. Johanna Tamminen (Finnish Meteorological Institute)
- Ongoing: Kernel-based regression for spectroscopy applications: prof. Heikki Haario and MSc Zina-Sabrina Duma (Lappeenranta University of Technology, Finland)
- Planned: Others (US universities)

- Technology transfer:
- Open sourcing

ESTO Earth Science Technology Office

14

14

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Publications

- Journal / Conference Papers
- Dissertations
- Other

List of Acronyms

• 1D	One-dimensional
• CMS	Content Management System.
• CSV	Comma Separated Variables
• DEM	Digital Elevation Model
• DS	Decadal Survey
• Etc.	

17

17

DTAS: A Prototype Digital Twin for Air-Sea Interactions

Alison Gray (PI, University of Washington)
Subhashis Hazarika, Ashesh Chattopadhyay
(Co-Is, Palo Alto Research Center)

AIST-21-0091 Annual Technical Review
June 23, 2023

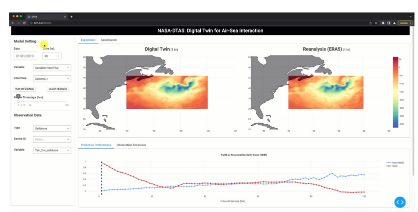
Team: Robert Drucker, Naomi Wharton (University of Washington)

DTAS: Digital Twin for Air-Sea Interactions

PI: Gray, Alison, University of Washington

Objective

- Develop a core hybrid physics-informed AI model for air-sea interactions: an essential part of the general circulation model as they help modulate the Earth's weather and climate.
- Assimilate novel data sources (Saildrone, Glider) along with remote sensing and Argo floats to produce a real-time "Digital Twin" simulation.
- Develop a visual analysis system to visualize model predictions, analyze relationships and perform what-if investigations.



Approach

- Preprocess and fuse datasets from multiple sources—correct for varying spatial and temporal resolution
- Develop a physics informed hybrid model that has a spatial model (U-net) and temporal model (Transformers) along with governing equations for oceanic and atmospheric variables: train using remote sensing, Argo and Glider data; validate using Saildrone datasets.
- Develop a visual interface using Dash to enable user-defined queries and simulations

Co-Is/Partners: Palo Alto Research Center, Inc.

Key Milestones

• Data preprocessing	11/22
• Data assimilation	01/23
• Prototype model architecture	05/23
• Visual Analysis System	10/23
• Model training completed	11/23
• Complete model integration	01/24

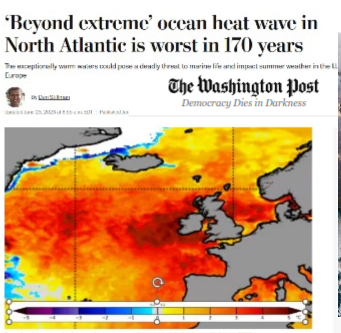
TRL_{in} = 2 **$TRL_{current}$ = 3**

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Background

Ocean-atmosphere interactions are crucial processes that impact many societally-important phenomena


'Beyond extreme' ocean heat wave in North Atlantic is worst in 170 years

The Washington Post
Democracy Dies in Darkness


https://cdn.star.nesdis.noaa.gov

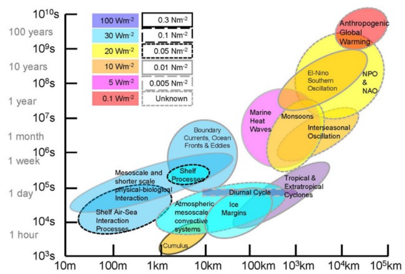


Climate Variability; Water & Energy; Atmospheric Comp; Earth Surface Interior; Weather; Eco Forecasting; Disasters; Health & Air Quality; Energy Management; Water & Food; Fires; Planetary Boundary Layer; Snow and Ice; etc.

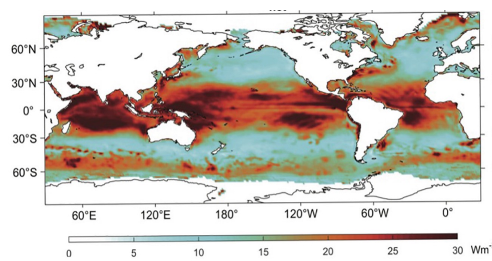
Background

However, these processes remain poorly constrained over vast areas of the planet

Flux Accuracies and Processes



Stommel diagram showing target flux accuracies



Standard deviation in annual mean net heat flux across 10 gridded flux estimates

Cronin et al. 2019

5

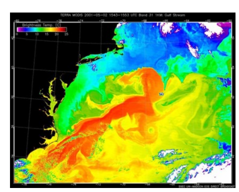
Objectives

Overall objective is to develop a *prototype* version of an **interpretable *Digital Twin for Air-Sea interactions*** that will advance **near-real-time estimation** of air-sea fluxes and their **uncertainties**.

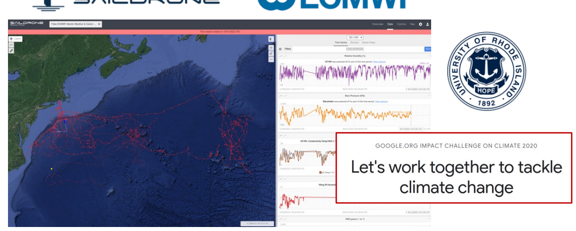
Four specific objectives

- Real-time, multi-scale data fusion
- Hybrid physics-informed AI model
- Framework for post-hoc analysis, including feature importance and uncertainty quantification
- Visual interface for model diagnosis and "what-if" investigations

Prototype will focus on turbulent heat fluxes in the Gulf Stream region of the North Atlantic.



SAILDRONE ECMWF



6

Presentation Contents

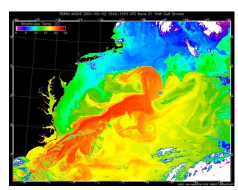
Technical and Science Advancements

Overall objective is to develop a *prototype* version of an **interpretable *Digital Twin for Air-Sea interactions*** that will advance **near-real-time estimation** of air-sea fluxes and their **uncertainties**.

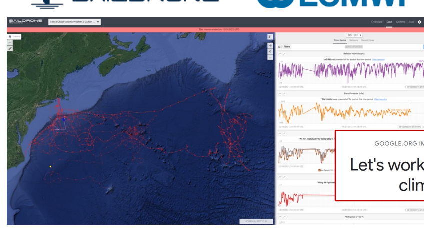
Four specific objectives

- Real-time, multi-scale data fusion
- Hybrid physics-informed AI model
- Framework for post-hoc analysis, including feature importance and uncertainty quantification
- Visual interface for model diagnosis and "what-if" investigations

Prototype will focus on turbulent heat fluxes in the Gulf Stream region of the North Atlantic.



SAILDRONE ECMWF

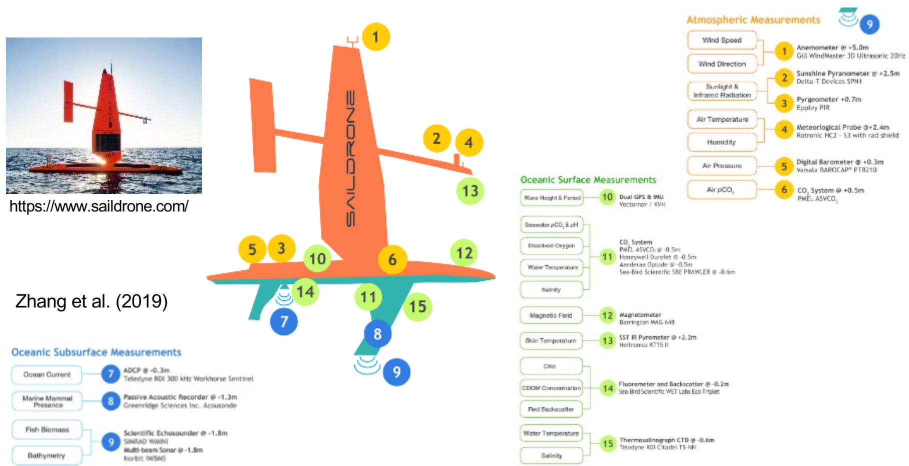


GOOGLE.ORG IMPACT CHALLENGE ON CLIMATE 2020

Let's work together to tackle climate change

Data Processing and Fusion

Saildrone autonomously collects simultaneous, high-frequency observations above and below the sea surface

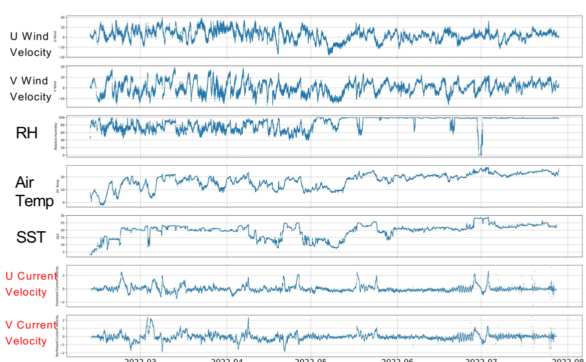


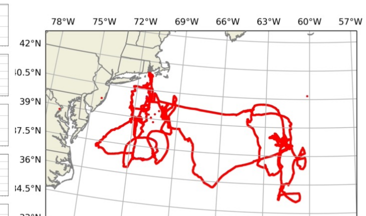
https://www.saildrone.com/

Zhang et al. (2019)

Data Processing and Fusion

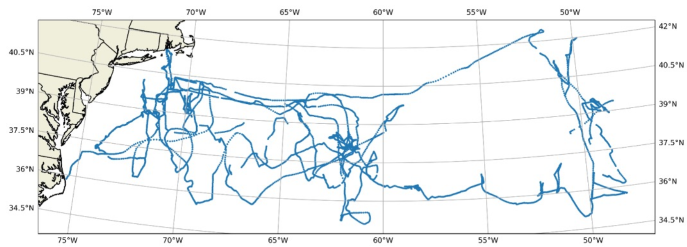
Saildrone Mission #1089: 02/11/2022 – 07/28/2022





Data Processing and Fusion

All Saildrone missions in Gulf Stream region



- Wind speed adjusted for ocean currents (m/s)
- Air temperature (ºC)
- Relative humidity (%)
- Sea surface temperature (ºC)

→ Latent heat flux (W/m^2)
Sensible heat flux (W/m^2)

The Coupled Ocean–Atmosphere Response Experiment (COARE) algorithm estimates heat fluxes from "bulk" state variables measured at the surface and some height within the surface layer:

$$Q_{lat} = \rho L_v \langle wq \rangle \approx \rho L_v C_E S \Delta Q$$
$$Q_{sen} = \rho c_p \langle w\theta \rangle \approx \rho c_p C_H S \Delta \Theta$$

ESTO Earth Science Technology Office

11

Data Processing and Fusion

Numerous relevant fields from reanalysis and L4 data products have been collocated with Saildrone data

DataSet	Description	Parameters	resolution	notes
ERA5	ECMWF atmospheric reanalysis	Q_sw, Q_lw Q_sensible, Q_latent Tair, SST, uwnd, vwnd	0.25 deg hourly mean	all parameters are 1-hour means 1/2 hour added to time coordinates[1] all heatfluxes are positive down
JRA	Japan Meteorological Agency (JMA) Japanese 55-year reanalysis	Q_sw, Q_sw_up, Q_sw_down Q_lw, Q_lw_up , Q_lw_down Q_sensible, Q_latent Q_radiative = Q_sw + Q_lw Q_turbulent = Q_sensible + Q_latent	0.56 deg 3-hr avg	all parameters are 3-hour means Times are 3-hr avg center points[1] Q_sw_up and Q_lw_up are positive up[4]; all other heatfluxes are positive down
CERES	NASA Clouds and the Earth's Radiant Energy System (CERES) Hourly gridded radiative surface fluxes	Q_sw_up, Q_sw_down Q_lw_up , Q_lw_down Q_sw = Q_sw_down - Q_sw_up Q_lw = Q_lw_down - Q_lw_up Q_radiative = Q_lw + Q_lw	1 deg hourly mean	all parameters are 1-hour means Times are 1-hr avg center points Q_sw_up and Q_lw_up are positive up[4]; all other heatfluxes are positive down
OAFlux	WHOI objectively analyzed air-sea fluxes	Q_sensible, Q_latent Q_turbulent = Q_sensible + Q_latent SST, Tair, wind_speed spec_hum, evap_rate	1 deg daily mean	all parameters are daily mean all heatfluxes are positive down Times are daily avg centers(12:00 UTC)
GHRSST_ REMSS	REMSS Global Hi-res SST Analysis	SST, SST_error, landmask	0.088 deg daily mean[2]	Times are the nominal time of analysis (12:00 UTC)
CCMP_ wind	REMSS Cross-Calibrated Multi-Platform surface wind velocity	uwnd, vwnd	0.25 deg 6-hourly	Times are the nominal time of analysis (0,6,12,18 UTC)[3]
CMEMS_ currents	Copernicus Marine Environment Monitoring Service 0.5m ocean currents	surface_current_u surface_current_v	1/12 deg 6-hourly	Times are the nominal time of analysis (0,6,12,18 UTC)[3] no data for 2019
CMEMS_ color	CMEMS ocean color	chlorophyl-a	daily	Times are daily avg centers(12:00 UTC)
CMEMS_ sealevel	SEALEVEL_GLO_PHY_ L4_NRT_OBSERVATIO NS_008_046	SLA, ADT geo_current_u, geo_current_v	0.25 deg daily	ADT is SSH above geoid SLA is SSH above mean sea surface ADT= SLA + mean dynamic topography no data for 2019
MERRA2	Modern-Era Retrospective analysis for Research and Applications (MERRA2)	Q_sw, Q_lw, Q_sensible, Q_latent uwnd, vwnd T_skin, T10m spec_hum rain_rate_kg	0.5 deg lat 0.625 deg lon hourly	all parameters are 1-hour means Times are 1-hr avg center points all heatfluxes are positive down

NOTES:

[1] ERA5 and JRA sample times are converted to the average center time.

[2] Not sure if GHRSST_REMSS are daily averages

[3] Not sure if CCMP_wind times are center or initial time of averages

[4] All heatfluxes except Q_sw_up and Q_lw_up converted to positive downward

Parameters:

SST = foundational sea surface temperature (°C)
T_skin = sea surface skin temperature (°C)
Tair = air temperature at 2m (°C)
T10m = air temperature at 10m (°C)
spec_hum = specific humidity ($kg\ kg^{-1}$)
rain_rate_kg = mass rain rate ($kg\ m^{-2}\ s^{-1}$)
uwnd = eastward wind component at 10m ($m\ s^{-1}$)
vwnd = northward wind component at 10m ($m\ s^{-1}$)
wind_speed = 10m wind speed ($m\ s^{-1}$)
surface_current_u = eastward 0.5m ocean current ($m\ s^{-1}$)
surface_current_v = northward 0.5m ocean current ($m\ s^{-1}$)
geo_current_u = eastward geostrophic ocean current ($m\ s^{-1}$)
geo_current_v = northward geostrophic ocean current ($m\ s^{-1}$)
SLA = Sea Level Anomaly (m)
ADT = Absolute Dynamic Topography (m)
chl_a = chlorophyl-a concentration ($mg\ m^{-3}$')
MLD = mixed layer depth (m)
T_{ML} = mixed layer temperature (°C)
S_{ML} = mixed layer salinity (PSU)
OHC = ocean heat content (J)

ESTO Earth Science Technology Office

12

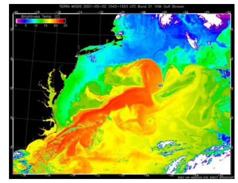
Technical and Science Advancements

Overall objective is to develop a *prototype* version of an **interpretable *Digital Twin for Air-Sea interactions*** that will advance **near-real-time estimation** of air-sea fluxes and their **uncertainties**.

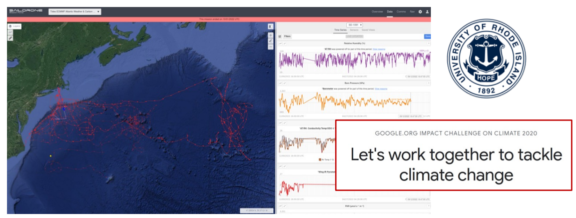
Four specific objectives

- Real-time, multi-scale data fusion
- Hybrid physics-informed AI model
- Framework for post-hoc analysis, including feature importance and uncertainty quantification
- Visual interface for model diagnosis and "what-if" investigations

Prototype will focus on <u>turbulent heat fluxes</u> in the Gulf Stream region of the North Atlantic.

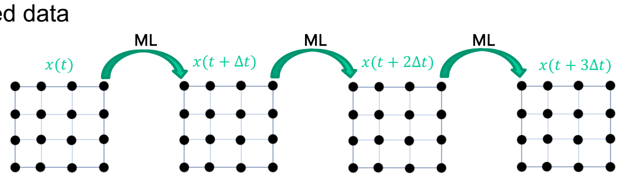


SAILDRONE ECMWF

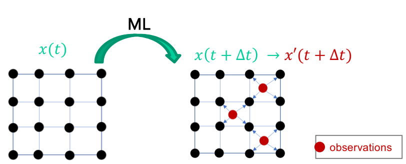


Physics-informed ML Model

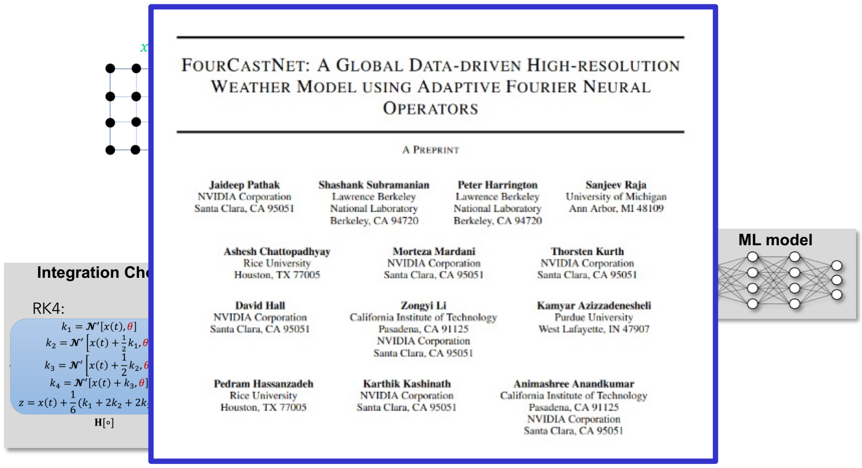
- Train a **Physics-informed ML model** to predict autoregressively on regular gridded data



- Perform **data assimilation (DA)** with combined Saildrone and L4 data products to improve model predictions

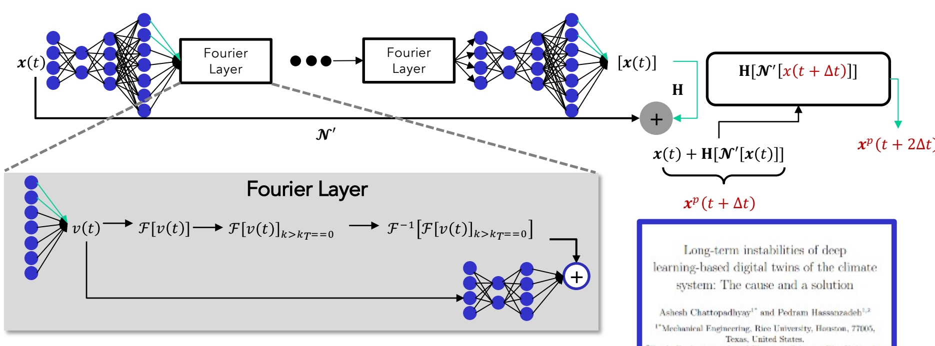


Physics-informed ML Model



Physics-informed ML Model

Model Architecture with underlying Neural Operators



Loss Function:

$$L(\theta)=\frac{1}{T}\sum_{t=0}^{T}\sum_{k=1}^{2}\|x^p(t+k\Delta t)-x^{true}(t+k\Delta t)\|_2+\|\hat{x}^p(t+k\Delta t)-\hat{x}^{true}(t+k\Delta t)\|_2$$

Physics-informed ML Model

Model Training Underway

- Choices marked in blue are the ones that we have trained models for

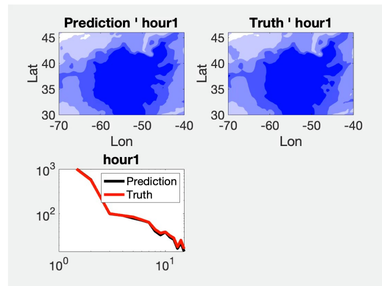
Several choices for $x(t)$:

1. $x(t) = (mslhf)$
2. $x(t) = (msshf)$
3. $x(t) = \begin{pmatrix} mslhf \\ msshf \end{pmatrix}$
4. $u(t) = \begin{pmatrix} mslhf \\ msshf \\ T2m \end{pmatrix}$
5. $u(t) = \begin{pmatrix} u10 \\ v10 \\ T2m \end{pmatrix}$

Several choices for Δt:

1. $\Delta t = 1hr$
2. $\Delta t = 3hr$
3. $\Delta t = 6hr$

34 hrs hourly prediction for *mslhf* variable along with spectral loss to track prediction stability:



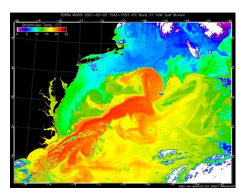
Technical and Science Advancements

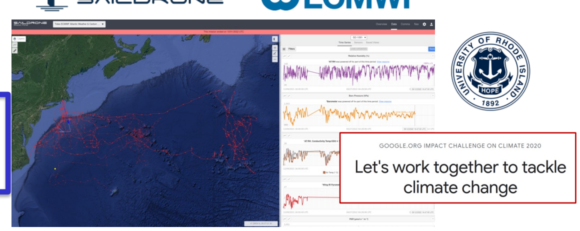
Overall objective is to develop a *prototype* version of an **interpretable *Digital Twin for Air-Sea interactions*** that will advance **near-real-time estimation** of air-sea fluxes and their **uncertainties**.

Four specific objectives

- Real-time, multi-scale data fusion
- Hybrid physics-informed AI model
- Framework for post-hoc analysis, including feature importance and uncertainty quantification
- Visual interface for model diagnosis and "what-if" investigations

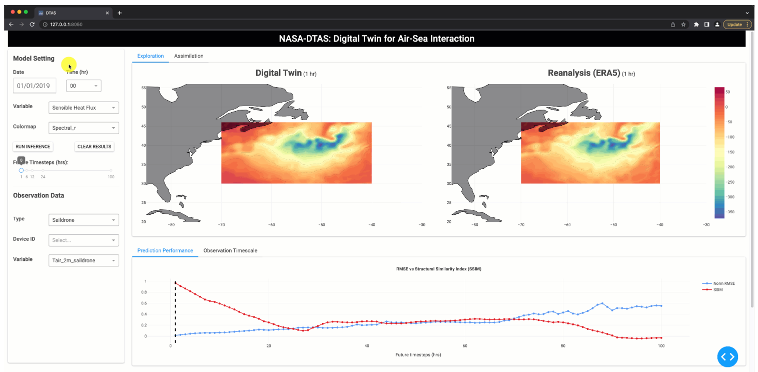
Prototype will focus on turbulent heat fluxes in the Gulf Stream region of the North Atlantic.





Visualization Interface

We have developed an initial prototype system to visualize the Digital Twin results and interact with Saildrone data

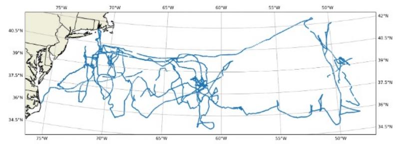


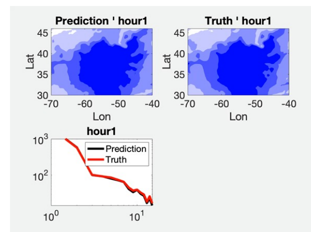
Presentation Contents

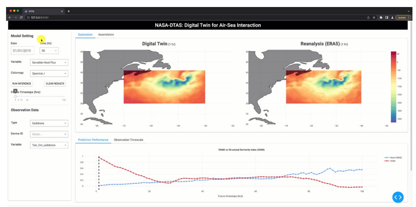
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Summary of Accomplishments

- Processed novel, in situ datasets and observational data products
- Developed physics-informed ML model based on FourCastNet architecture
- Created initial visual interface for interacting with digital twin

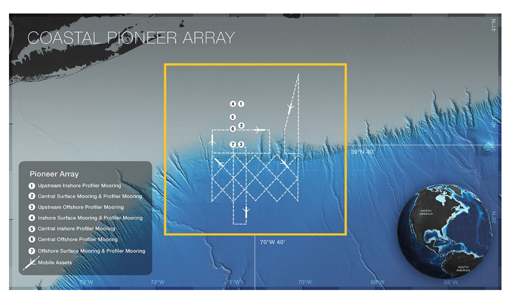






Future Plans

- Test various choices for ML model formulation
- Add multi-dimensional data assimilation of observational datasets through Saildrone measurements
- Process independent flux dataset for validation
- Develop framework for post-hoc analysis, including uncertainty quantification
- Refine visual interface to add capacity for model diagnosis and "what-if" investigations



Presentation Contents

Actual or Potential Infusions and Collaborations

- NASA Decadal Survey Incubation Team - PI Gray has become a member of the NASA Planetary Boundary Layer Incubation Team (Technology co-leads: Jeff Piepmeier & Amin Nehrir, NASA ; Science lead: Carol Anne Clayson, WHOI)
- Collaboration with FourCastNet development group via new team member Ashesh Chattopadhyay

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications

Publications

Gray, A.R. DTAS: A Prototype Digital Twin for Air-Sea interactions. Earth System Digital Twins (ESDT) Workshop, October 2022, Washington D.C.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K. and Hassanzadeh, P., 2022. Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214.

Chattopadhyay, A. and Hassanzadeh, P., 2023. Long-term instabilities of deep learning-based digital twins of the climate system: The cause and a solution. arXiv preprint arXiv:2304.07029.

Predicting What We Breathe

Jeanne Holm (PI, Deputy Mayor, City of Los Angeles)
Mohammad Pourhomayoun (Co-I, California State University, Los Angeles)
Dawn Comer (Project Manager, City of Los Angeles)
Chris Hagerbaumer (Co-I, OpenAQ)

AIST-18-0099 Report
June 23, 2023

1

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications
- Appendix

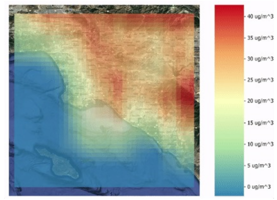
2

Predicting What We Breathe

PI: Jeanne Holm, City of Los Angeles

Objective

- Increase the accessibility and use of space data by using machine learning to help cities predict air quality (AQ) in ways that can be acted upon to improve human health outcomes.
- Increase the use of space data, with an additional scope on the intersection of health issues (with a focus on air quality) and transportation, particularly the new operational planning for urban air mobility.
- Provide these tools and algorithms to future Earth science missions (e.g., MAIA) to provide rapid ground truth, combine multiple data sources, and support more rapid use of mission data.



Predicted Air Pollution

Approach

- Develop machine learning (ML) algorithms to create predictive models for air quality based on measurements of 2.5 micron particulate matter ($PM_{2.5}$), Ozone, NOx and other air pollutants.
- Develop a big data analytics algorithm for integrating ground and space data..
- Develop predictive models for health risk using deep learning and machine learning.
- Build $PM_{2.5}$ stack for integrating ground and space data.
- Create a model for cities with shared attributes to understand predictions and effective interventions.
- Implement models for megacities: Los Angeles, Mexico City, Durban (South Africa), and London
- Developed the Digital Twin models for transportation and air quality

Co-Is/Partners: M. Pourhomayoun, Cal State LA; C. Hagerbaumer, OpenAQ; AQMD

Key Milestones

- Identify AQ data data for modeling 06/20
- Identify ML model 07/20
- Develop ML algorithm 12/20
- Identify City interventions 11/20
- Determine megacity attributes 08/20
- Complete ML training runs 09/20
- Validate algorithm 10/21
- Extend ML algorithm to additional sensor data 07/22
- Complete Pre- and post-intervention analysis 09/22
- 5 predictive models for predicting PM2.5, NO2, O3, NO, CO are developed and transferred to AWS to run on the cloud 12/22
- Formalize data sharing agreements, selection of datasets, and parameters for digital twin 01/23
- Identify data models for urban air mobility with FAA, LAWA, and LADOT 03/23
- First digital twin models for Transportation and Air Pollution 06/23

TRL_{in} = 2 **$TRL_{current}$ = 4**

Current State

- **Administrative**
  - Contracts established between the City and OpenAQ and Cal State L.A.
  - Participated in ESIP meetings, MAIA early adopter meetings, and AIST Annual Reviews
  - Launched project website - airquality.lacity.org and email address - airquality@lacity.org
  - Project Augmentation approved for Digital Twin for Transportation and Air Quality
- **Data Preparation**
  - Additional ground-based sensor data to the model including low cost sensors
  - Prepared the data for other cities, feedback from other cities
  - Baselined initial digital twin idea for AQ in Los Angeles
  - Identified ground-based and satellite datasets available from NASA, OpenAQ, and existing City department projects
  - Established regular engagement within the AQ data community to collaborate on best practices for accessing and using data (NASA, OpenAQ, L.A. County Health, etc.)
  - Used NASA satellite data for machine learning algorithms
  - Identify new data with FAA, LAWA, and LADOT
- **Technical Preparation**
  - Data processing and integration
  - Designing machine learning approaches
  - Developing and training machine learning Algorithms for discovering spatiotemporal patterns in the data and make predictions for PM2.5, NO2, O3, NO, CO
  - ESIP Evaluation completed
  - First digital twin models for Transportation and Air Pollution

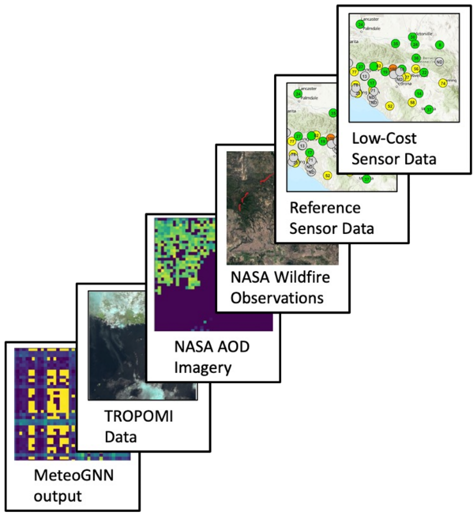
Current State (continued)

- **Community Engagement**
  - International
    - Held workshop at International Astronautical Congress in Dubai, U.A.E. with European Space Agency, African Space Union, South African Space Agency, and Rwandan Space Agency for data and support to African sister cities (October 2021)
    - AQ sister cities attended International workshop (Feb 2022)
      - Mexico City, Durban and London used the PWWB algorithm
    - Published and presented 18 peer-reviewed papers and 3 meeting papers
  - Regional
    - Continued engagement with community advocates (Anthem Blue Cross, Southern California Asthma Association, SmartAirLA, Pacoima Beautiful, City Council, and AQMD)
    - Augmented reality app with Agents of Discovery for community science
      - 4 Air Quality based Missions launched June 2022
      - 1 AQ based mission launched April 2023
    - Defined citizen science project with LA Public Library and SafeCast sensors
    - Identified AQ interventions to measure
    - Provided sensors to community-based organizations, conducted sessions with community members and Purple Air trainers
    - Engaged with the L.A. Public Library and the LA Mayor's Youth Council for Climate Action to increase community awareness of our Agents of Discovery Air Quality missions
    - Partnered with Agents of Discovery and Wonders of Wildlife to provide an educational video on Predicting Air Quality

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications
- Appendix

Predictive Model and Data



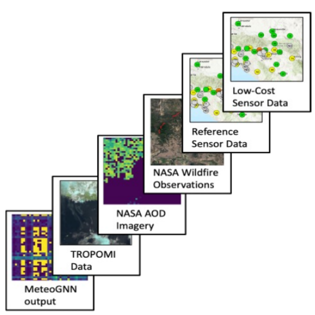
New Dataset added
- NASA AOD Imagery
- ESA/NASA TROPOMI Data
- NASA Wildfire data: MISR, MODIS, FRP
- Reference Sensors
- Low Cost Sensors: 32 AQMD and 48 PurpleAir community maintained sensors.
- Meteorological data
- LandSat products

Data Processing and Data Fusion
- Preprocessing and cleansing
- Outliers/trustworthiness and missing values
- Feature extraction and knowledge discovery
- Feature selection and dimensionality Reduction
- Format matching and alignments

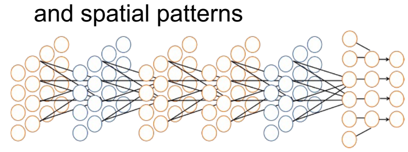
7

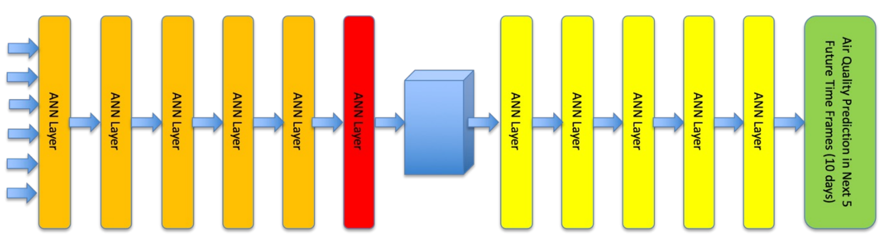
7

Deep Neural Networks for Predictive Models



- Developed 12 Predictive models. The predictive models include several very deep neural nets to capture and learn both temporal and spatial patterns

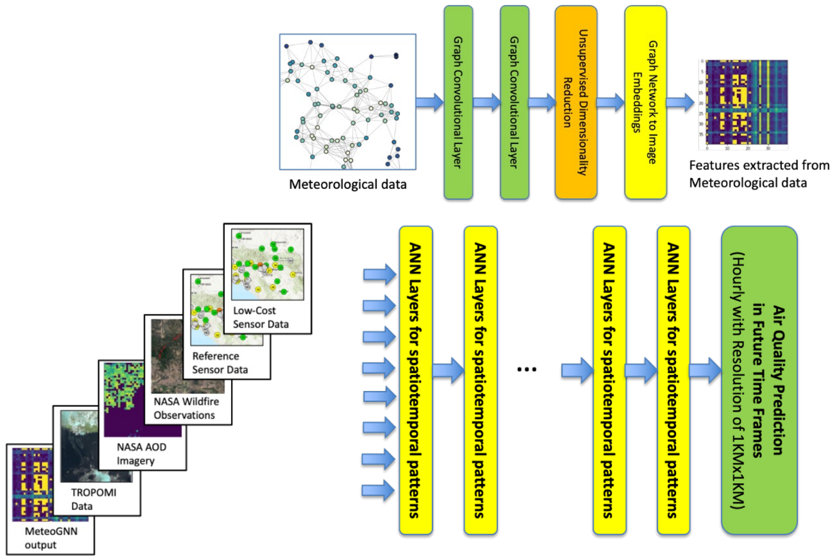




8

8

Deep Neural Networks for Predictive Models

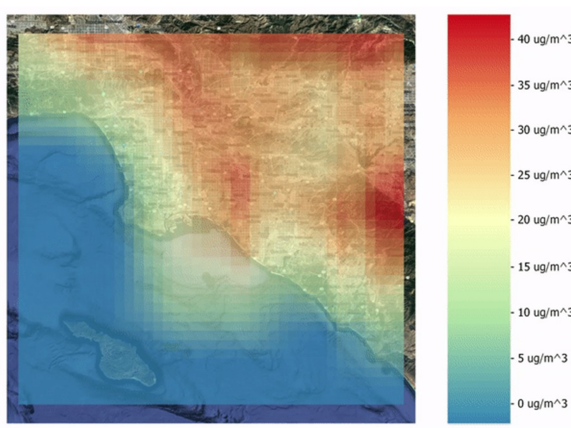


ESTO Earth Science Technology Office

9

Prediction Results

The average accuracy for 24-hour prediction over all site locations in LA County is 94.56%.



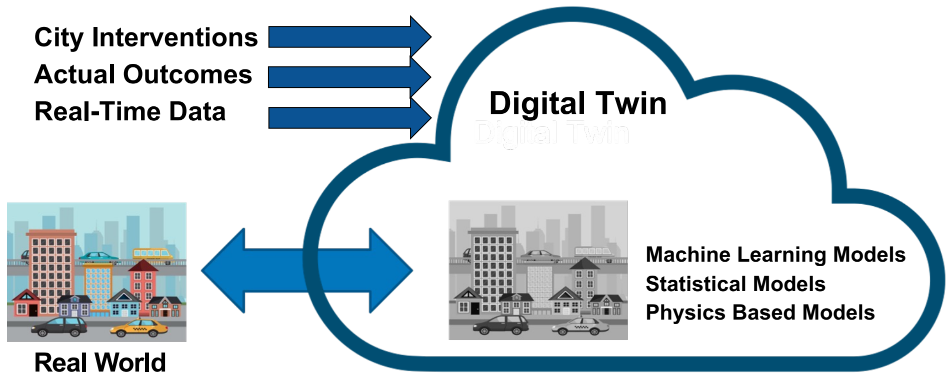
- **Temporal Resolution: hourly prediction**
- **Spatial Resolution: 250 m^2 (previously 1 km^2)**

ESTO Earth Science Technology Office

10

Digital Twin for Air Pollution and Transportation

We have developed a digital twin including AI models, data analytics and ML algorithms, statistical models, and data visualization to replicate the real-world system and its behavior, and to understand and predict urban air pollution.

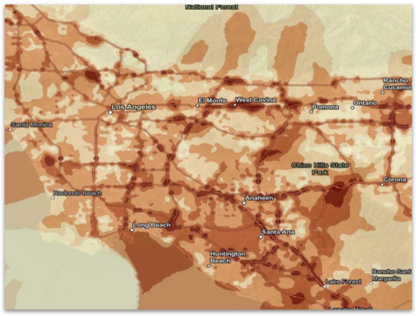


Digital Twin for Air Pollution and Transportation

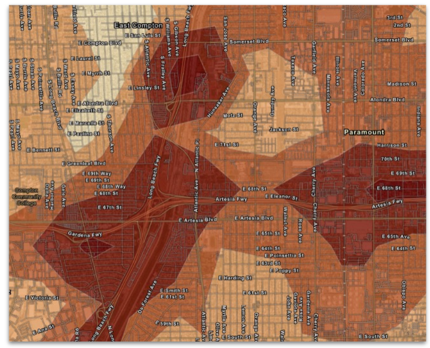
- The digital twin that **visualizes AQ predictions** for various types of air pollutants on various time scales
- The digital twin visualizes and **connects real-time air quality to transportation**.
- Through this work we will be able to visualize and **model what-if scenarios** showing the impact of changing modes of transportation or modifying public transportation on air quality.
- More importantly, to understand the impact of adding urban **air mobility** on transportation and air quality.

Digital Twin for Air Pollution and Transportation

The impact of Fossil Fuel Combustion by Cars, Trucks, and Busses

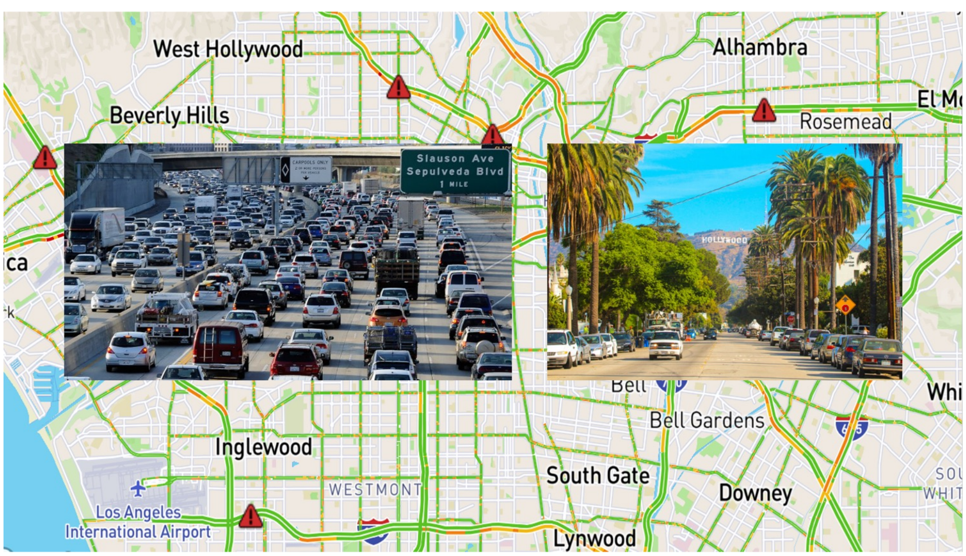


The projected air pollution created by Fossil Fuel Combustion by Cars, Trucks, and Busses



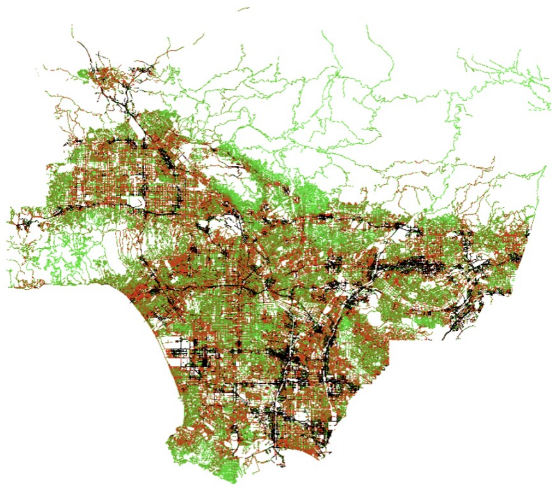
ESTO Earth Science Technology Office

Digital Twin for Air Pollution and Transportation



ESTO Earth Science Technology Office

Digital Twin for Air Pollution and Transportation



- Compute live traffic counts in real-time for every location.
- Convert traffic counts into live traffic-driven air pollution.

- Real-time Traffic Flow
- Static Traffic Count
- Real-time Traffic Count

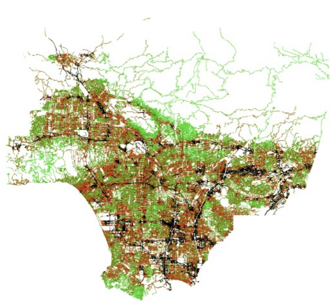
- Need to consider car speed and type of cars in each street

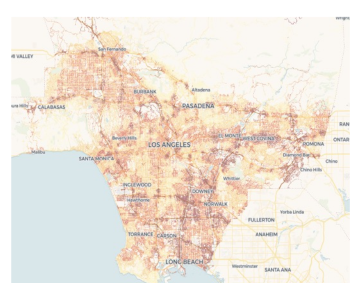
ESTO Earth Science Technology Office

15

Digital Twin Model for Air Quality and Transportation

Real-time Traffic Flow → Statistical modeling for every street in the entire city → Real-time Traffic Count → Vehicle Fuel Combustion Information → Real-time Transportation Air Pollution

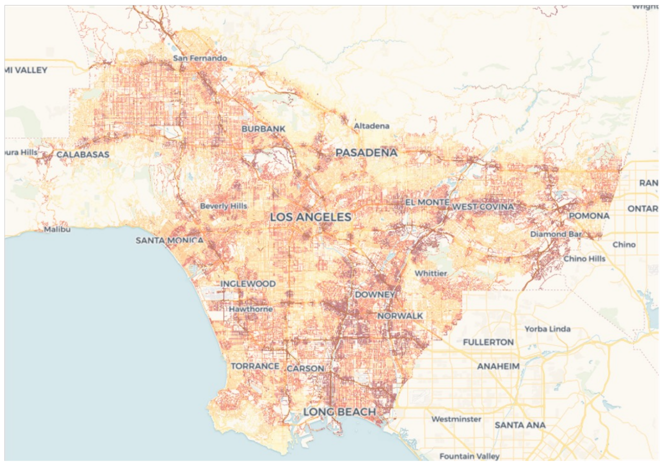




ESTO Earth Science Technology Office

16

Digital Twin Model for Air Quality and Transportation

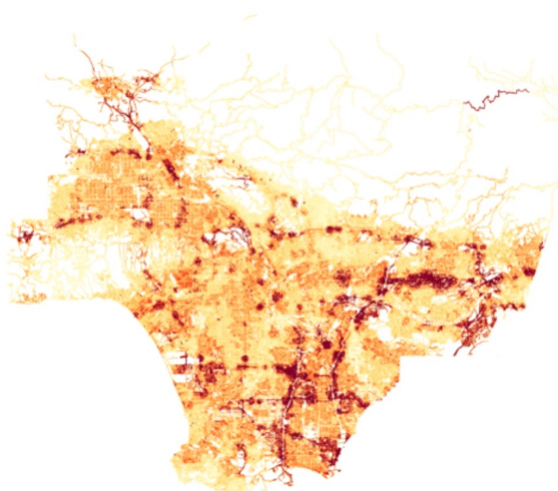


Real-time PM2.5
created by vehicles
per second

ESTO Earth Science Technology Office

17

Digital Twin Model for Air Quality and Transportation

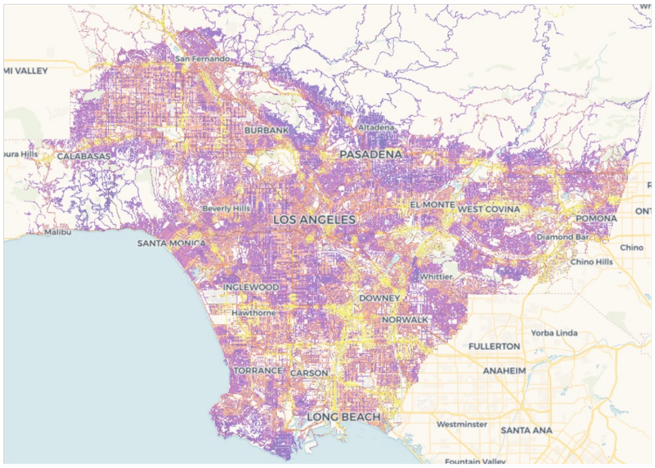


Real-time PM2.5
created by vehicles
per second

ESTO Earth Science Technology Office

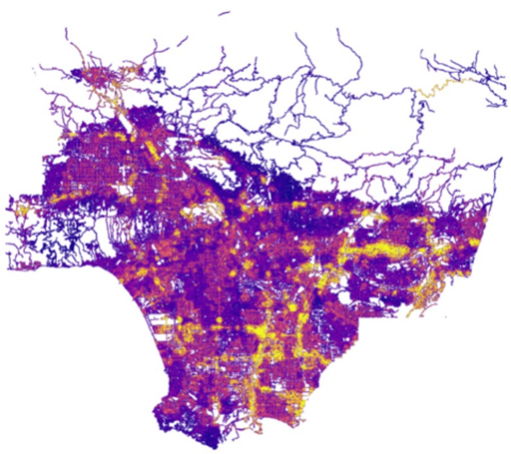
18

Digital Twin Model for Air Quality and Transportation



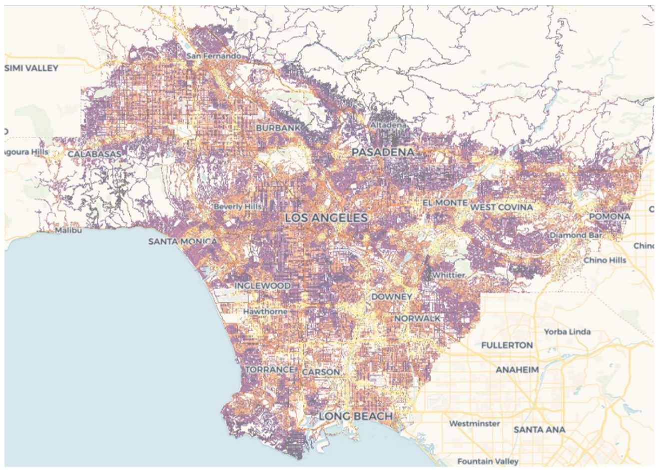
Real-time NO2 created by vehicles per second

Digital Twin Model for Air Quality and Transportation



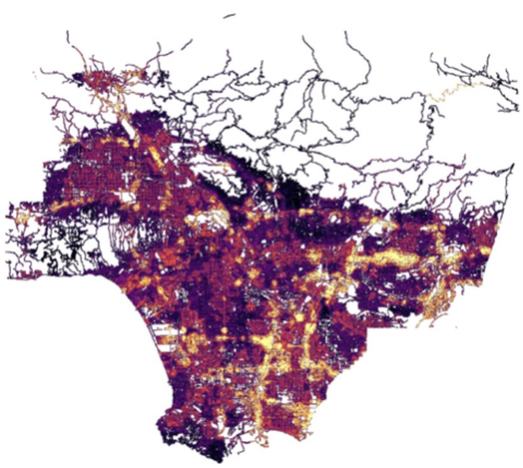
Real-time NO2 created by vehicles per second

Digital Twin Model for Air Quality and Transportation



Real-time CO2 created by vehicles per second

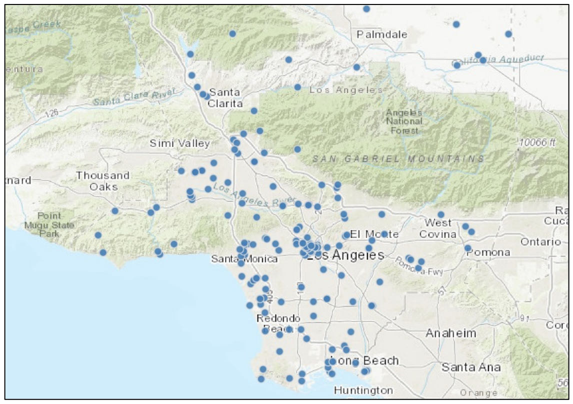
Digital Twin Model for Air Quality and Transportation



Real-time CO2 created by vehicles per second

Digital Twin for Air Pollution and Transportation

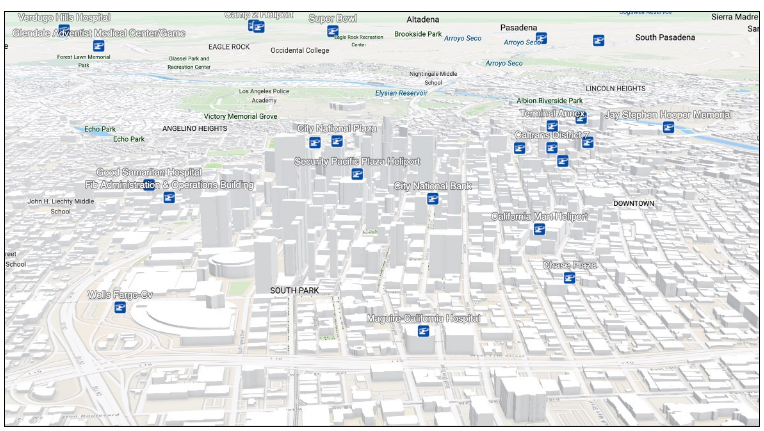
- Next Steps: Understand the impact of adding urban air mobility on transportation and AQ
- Digital Twin expansion to include air mobility based on city regulations
- Locations of Heliports in Los Angeles County



ESTO Earth Science Technology Office

Digital Twin for Air Pollution and Transportation

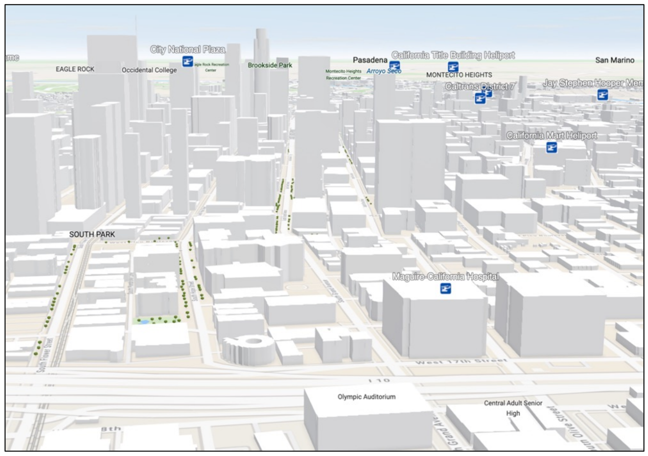
- Locations of Heliports in Los Angeles County



ESTO Earth Science Technology Office

Digital Twin for Air Pollution and Transportation

- Locations of Heliports in Los Angeles County



Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications
- Appendix

Plan Forward

Next Steps

PWWB
- Developing more predictive models for predicting other air pollutants
- Adding new datasets to the predictive models including more high-resolution satellite observations from NASA and new fire/smoke satellite data as well as more sensor data.
- Continue evolution of model, algorithms, and validation
- Continue to align augmented project scope of Developing an ACF for Air Quality as a collaboration among identified AIST projects - PWWB model has been merged into ACF system

Transition to Global Digital Twin for Transportation and Air Quality
- Developing more robust digital twin for L.A. and model with other global cities
- Continue to identify and integrate local data (health, polluters, traffic, roads, ports) from IOT and in-situ sensors; this includes the addition of data from the FAA, the City, NASA, and urban air manufacturers to run scenarios on the impact of moving traffic to a three dimensional space.
- Launch augmented project scope on Creating A City Global Digital Twin
- Understand the impact of adding urban air mobility on transportation and AQ
- Continue developing Digital Twin models to include air mobility based on city regulations
- Continue to engage community for environmental justice for awareness and support, and healthcare partners (Propeller Health, Anthem Blue Cross, Southern California Asthma Association) to improve health outcomes

27

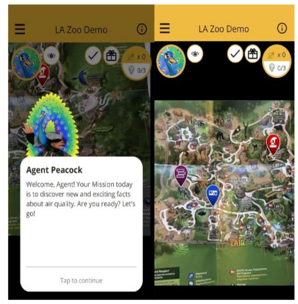
Presentation Contents

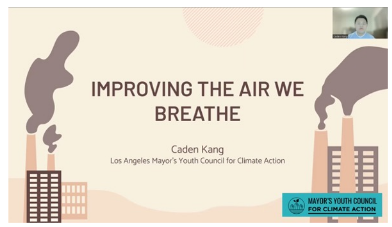
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications
- Appendix

28

Infusions and Collaborations

- Predicting What We Breathe Community Engagement
  - Distributed sensors in environmental justice communities
  - Regional community workshop (November 2021)
  - Partnered with LAUSD on sensor lessons
  - Partnered with Agents of Discovery for educational, community-based science missions (interactive learning experience as a part of extensive SoCal Explorer missions)
    - Localized City of L.A. experience
      - 3 Missions with 10 Challenges at Rec and Parks Locations
      - 1 Mission at L.A. Public Library
      - New Mission added for LA County Park (Spring 2023)
    - LA Mayor's Youth Council, Agents of Discovery and Wonders of Wildlife to present educational video on Air Quality
- Digital Twin - Integration of data from the FAA, the City, NASA, and urban air manufacturers to run scenarios on the impact of moving traffic to a three dimensional space.





Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Publications
- Appendix

Publications

Journal and Conference Papers: **18 peer-reviewed papers**, six meeting papers

- Paper published in ICDATA 2023:
  - "Predicting Atmospheric Air Pollution: A Convolutional-Transformer Approach for Spatial and Temporal Analysis of PM2.5"
- Paper published in CSCI 2022:
  - Multi-Pollutant Ground-Level Air Pollution Prediction through Deep MeteoGCN-ConvLSTM
- Paper published in ICDATA 2022:
  - "High-Resolution Spatiotemporal PM2.5 Prediction with Deep Convolutional LSTM using Atmospheric and Ground-level Data"
- Published in the Journal of Atmosphere 2022
  - Air Pollution Prediction through Deep Learning Using Multisource Meteorological, Wildfire, and Heat Data (https://www.mdpi.com/2073-4433/13/5/822)
- Two papers at the AGU (American Geophysical Union) Fall Meeting 2022
  - Global High-Resolution PM2.5 Prediction Applying Multi-source Big Data through Deep Convolutional LSTM
  - Digital Twin Cities for Air Quality Simulation

Publications (continued)

- Two papers published/presented in International Astronautical Congress (Sep 2022)
  - Federating Space, Air, and Ground Data for Air Quality Around the World
  - Creating Global Digital Twins to Improve Air Quality and COVID Outcomes
- Published in the journal of Air Quality, Atmosphere, and Health 2021
  - Predicting PM2.5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data (link.springer.com/article/10.1007/s11869-021-01126-3)
- Three papers at the AGU (American Geophysical Union) Fall Meeting 2021
  - Creating Global Digital Twins to Improve Air Quality and COVID Outcomes
  - Predicting PM2.5 Air Pollution using Deep Learning with Multisource Satellite and Ground-based Observations and Meteorological and Wildfire Big Data
  - AQACF: A Platform for Air Quality Analysis, Visualization, and Prediction
- Paper published in ICDATA 2021:
  - "A Comprehensive Analysis of Air Pollution and Equity During COVID-19"
- Two papers published/presented in 2020 International Conference on Computational Science and Computational Intelligence (CSCI'20: December 16-18, 2020, Las Vegas, USA) www.american-cse.org/csci2020/
  - *Satellite Image Atmospheric Air Pollution Prediction through Meteorological Graph Convolutional Network with Deep Convolutional LSTM*
  - *Sensor-Based Air Pollution Prediction Using Deep CNN-LSTM*

Publications (continued)

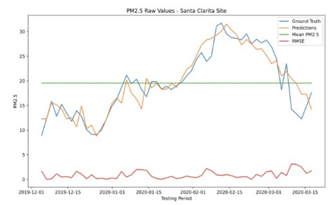
- Paper presented at ICDATA conference (July 27, 2020): presentation video
  - *Spatiotemporal Air Pollution Prediction with Deep ConvLSTM via Satellite Image Analysis*
- Presented project at the Environmental Law Institute (July 29, 2020)
  - ELI is supporting the U.S. EPA to characterize and learn from how states, tribes, and local governments are using citizen science in their programs
- Peer-reviewed paper at International Astronautical Congress (October 11, 2020)
- European Space Agency's Space for Twin Cities broadcast (November 2020)
- Workshops
  - International City Engagement Workshop (February 24, 2022)
  - Los Angeles Community Engagement Workshop (May 27, 2021)
- Two abstracts presented in AGU (American Geophysical Union) Fall Meeting (December2020)
  - *Particulate Matter Forecasting in Los Angeles County with Ground-Based Sensor Data Analytics*
  - *Real-Time Spatiotemporal NO2 Air Pollution Prediction with Deep Convolutional LSTM through Satellite Image Analytics*
- Other
  - Presented at LADWP-MWD University Symposium (May 19, 2022)
  - Project mentioned by Mayor Garcetti: SCAQMD Environmental Justice Conference
  - UN International Day of Clean Air: City of L.A. Social Media (September 7, 2020)
  - Clean Air Day - City of L.A.: Press Release and Social Media (October 7, 2020)
  - Presented at City of L.A. Chief Sustainability Officer Meeting (November 18, 2020)

Presentation Contents

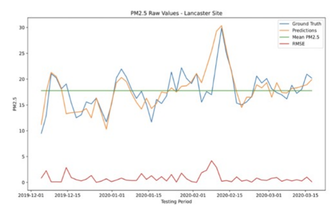
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Publications
- Appendix

Predicting PM2.5 Based on Satellite Observations, Ground Sensors, Meteorological Data, and Wildfire/Smoke Data

Santa Clarita Site PM 2.5 Observed Sensor Data vs Predicted



Lancaster Site PM 2.5 Observed Sensor Data vs Predicted



48-hour prediction Accuracy	Sensor Location
94%	Downtown LA
95%	Long Beach
91%	Lancaster
91%	Glendora
93%	Santa Clarita
93%	Reseda
95%	Long Beach – Rt 710

Predicting PM2.5 Based on Satellite Observations, Ground Sensors, Meteorological Data, and Wildfire/Smoke Data

Input data

- Satellite observations NASA MODIS
- Ground-based sensors (13 in L.A. County), hourly
- Wildfire/Smoke data from NASA MODIS, MERRA-2
- Meteorological data

10-Day Prediction Accuracy	Days
93%	2 days in future
90%	4 days in future
88%	6 days in future
83%	8 days in future
80%	10 days in future

Predicting Ozone Based on Satellite Observations, Ground Sensors, Meteorological Data, and Wildfire/Smoke Data

48-hour prediction Accuracy	Sensor Location
93.53%	Downtown LA
95.90%	Long Beach
91.25%	Santa Clarita
88.19%	Reseda
86.23%	Lancaster
87.35%	Glendora
91.45%	Westchester
87.49%	Pico Rivera
90.04%	Compton
92.87%	Pasadena
93.10%	West LA
92.13%	Azusa
90.59%	Pomona

Accuracy	Frame #
91%	Frame 1: 2 days in future
89%	Frame 2: 4 days in future
86%	Frame 3: 6 days in future
84%	Frame 4: 8 days in future
80%	Frame 5: 10 days in future

- Satellite observations NASA MODIS
- Ground-based sensors (13 in L.A. County), hourly
- Wildfire/Smoke data from NASA MODIS, MERRA-2
- Meteorological data

ESTO Earth Science Technology Office

Predicting NO$_2$ Based on Satellite Observations, Ground Sensors, Meteorological Data, and Wildfire/Smoke Data

24-hour prediction Accuracy	Sensor Location
93%	Downtown LA
91%	Long Beach
91%	Santa Clarita
89%	Reseda
87%	Lancaster
88%	Glendora
91%	Westchester
91%	Pico Rivera
95%	Compton
92%	Pasadena
90%	West LA
92%	Azusa
92%	Pomona

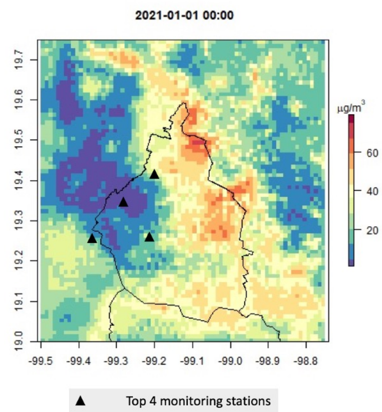
Accuracy	Frame #
87.62%	Frame 1: 2 days in future
84.15%	Frame 2: 4 days in future
82.38%	Frame 3: 6 days in future
79.06%	Frame 4: 8 days in future
72%	Frame 5: 10 days in future

- Satellite observations NASA MODIS
- Ground-based sensors (13 in L.A. County), hourly
- Wildfire/Smoke data from NASA MODIS, MERRA-2
- Meteorological data

ESTO Earth Science Technology Office

GOBIERNO DE LA CIUDAD DE MÉXICO | SECRETARÍA DEL MEDIO AMBIENTE | CIUDAD INNOVADORA Y DE DERECHOS

PWWB forecast in Mexico City for $PM_{2.5}$



Hourly predictions

- January 1st, 2021 - January 2nd, 2021
- 75km x 75km area of Mexico City
- Spatial resolution of 1 pixel per square kilometer

GOBIERNO DE LA CIUDAD DE MÉXICO | SECRETARÍA DEL MEDIO AMBIENTE | CIUDAD INNOVADORA Y DE DERECHOS

PWWB forecast in Mexico City for $PM_{2.5}$

Model accuracy

- November 1st, 2020 - December 31st, 2020
- 15 sensor locations
- Compared to ground truth data

Average Accuracy	Monitoring Station
94.87%	**Santa Fe**
94.72%	**Ajusco Medio***
94.53%	**Miguel Hidalgo***
94.51%	**Investigaciones Nucleares**
94.42%	Hospital General de México*
94.39%	Benito Juárez
94.34%	Tlalnepantla
94.19%	San Agustín
94.18%	Merced
94.17%	Gustavo A. Madero
93.95%	Ajusco
93.69%	Nezahualcóyotl
93.68%	Centro de Ciencias de la Atmósfera
93.56%	Xalostoc*
93.39%	UAM Xochimilco

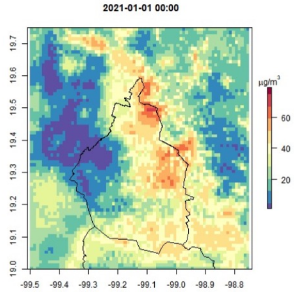
GOBIERNO DE LA CIUDAD DE MÉXICO | SECRETARÍA DEL MEDIO AMBIENTE

CIUDAD INNOVADORA Y DE DERECHOS

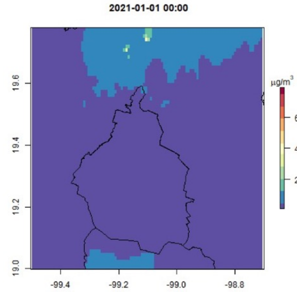
Comparison with Mexico City's current forecast model

$PM_{2.5}$ hourly predicted values

- PWWB

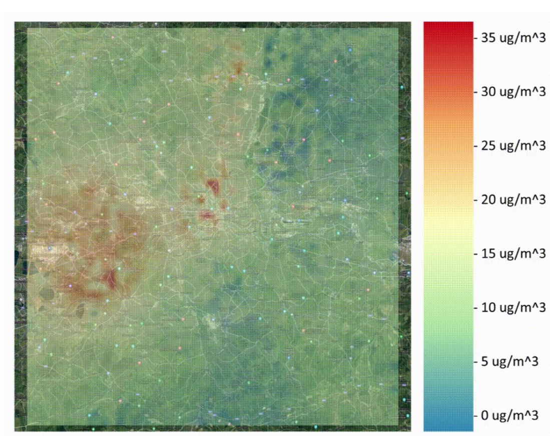


- SEDEMA 18z



London, U.K. PWWB Instance

Hourly prediction for Jan .1, 2022



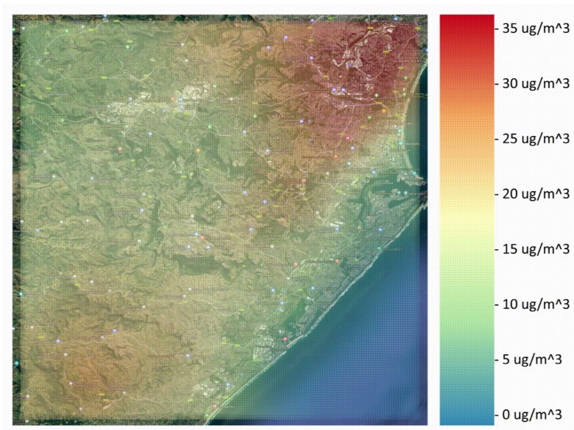
London Predicted PM2.5 Results

Predicted ground-level PM2.5 in 12 sensor locations in London hourly

Average Accuracy	Sensor Location
91.03%	London Teddington Bushy Park
91.14%	Kensington and Chelsea
91.04%	Sutton - Beddington Lane
89.61%	Camden - Bloomsbury
91.17%	City of London - Farringdon Street
90.90%	City of London - The Aldgate School
90.77%	Tower Hamlets - Blackwall
90.65%	Greenwich - Westhorne Avenue
91.03%	Greenwich - A206 Burrage Grove
90.92%	Greenwich - Plumstead High Street
90.98%	Greenwich - Falconwood FDMS
91.02%	Havering - Rainham

Durban, South Africa PWWB Instance

Hourly prediction for Jan. 1, 2022



Durban Predicted PM2.5 Results

Predicted ground-level PM2.5 in 4 sensor locations in Durban hourly

Average Accuracy	Sensor Location
94.54%	Settlers
94.53%	Wentworth
93.88%	Durban City Hall
95.61%	New Germany

AIST IDEAS:
Integrated Digital Earth Analysis System for Hydrology

Mid-year Review

Thomas Huang, Cedric David, Gary Doran, Jason Kang, Grace Llewellyn, Kevin Marlis,
Stepheny K. Perez (Dev Lead), Wai (William) Phyo, Joe T. Roberts,
Catalina Oaida Taglialatela, Megan Bull (Intern)
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

Earth Science Technology Office

IDEAS: Integrated Digital Earth Analysis System

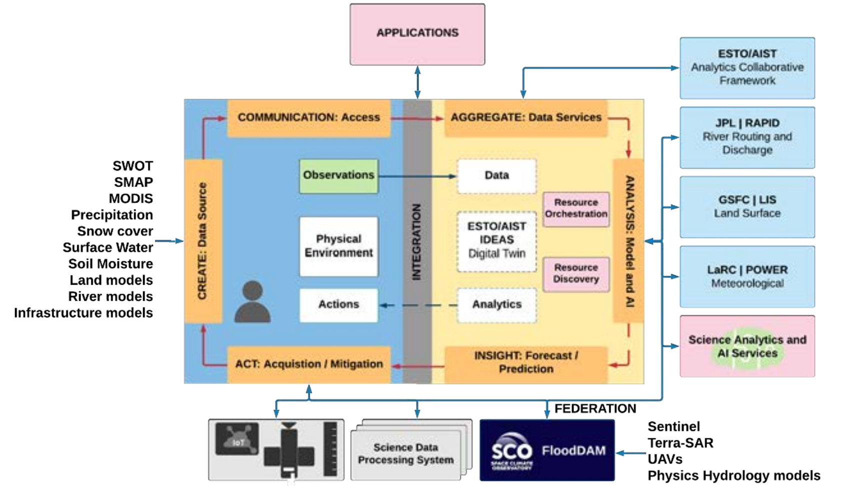
Thomas Huang, JPL

Objective

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 is a Digital Twin architecture enabling analysis of interacting Earth systems, such as for floods and their impacts, from cross-mission models and data sets.

Approach

Architecture, interfaces and workflows. Build on SDAP, an AIST-funded architecture for collaborative analysis across data sources. Enable coarse coordination across models by developing analytic workflows that orchestrate model execution and open interfaces for exchanging geophysical variables among models.

New Observing Strategies. Dynamic data assimilation.

Integrate models and data sets for interacting hydrologic and human systems, including the Land Information System (LIS), RAPID (rivers), and POWER (power and agroclimatology).

Federate with SCO-FloodDAM physics-based flood models, datasets and image generation.

Co-Is/Partners: S. Kumar, GSFC; P. Stackhouse, D. Borges, LaRC; C. David, C. Oaida, JPL; S. Baillarin, CNES/SCO (FloodDAM)

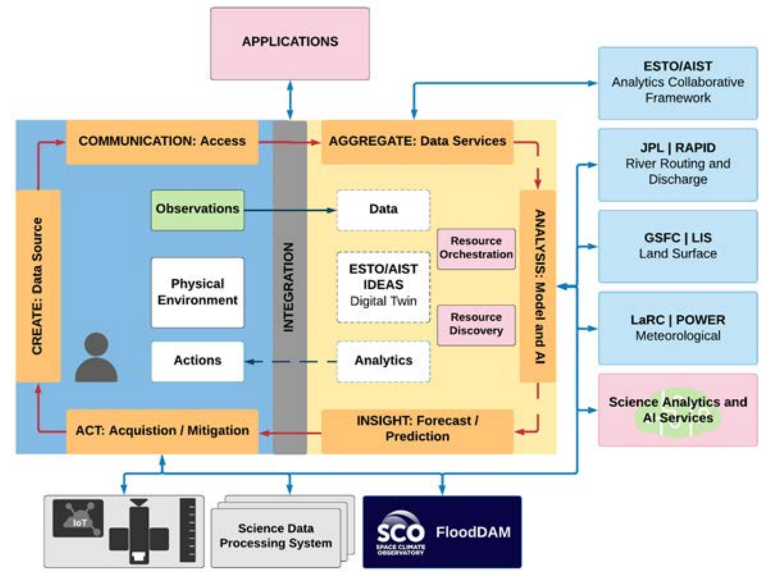
Key Milestones

- Architecture design. Identification of specific datasets, models, and analysis use cases. — 03/22
- Initial prototype of IDEAS system. Integration of data sets and interfaces with models. — 06/22
- IDEAS demonstration: analysis across models and data sets of current Earth system state. — 09/22
- Integration of ML models for event recognition and prediction — 03/23
- Federation with SCO-FloodDAM models and data. — 06/23
- IDEAS demonstration: forecasts and what-if projections — 09/23

$TRL_{in} = 3$ $TRL_{current} = 3$

ESTO/AIST Integrated Digital Earth Analysis System (IDEAS)

- **IDEAS -** NASA ESTO/AIST 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
- 2-year project using water cycle and flood analysis as the prototype application for integrated data and science
- Multi-Agency and Multi-Center partnership (JPL, GSFC, LaRC, CNES)
- Integrate 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
  - **Space for Climate Observatory (SCO) FloodDAM and CNES:** Automated service to reliably detect, monitor and assess flood events globally
- 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)



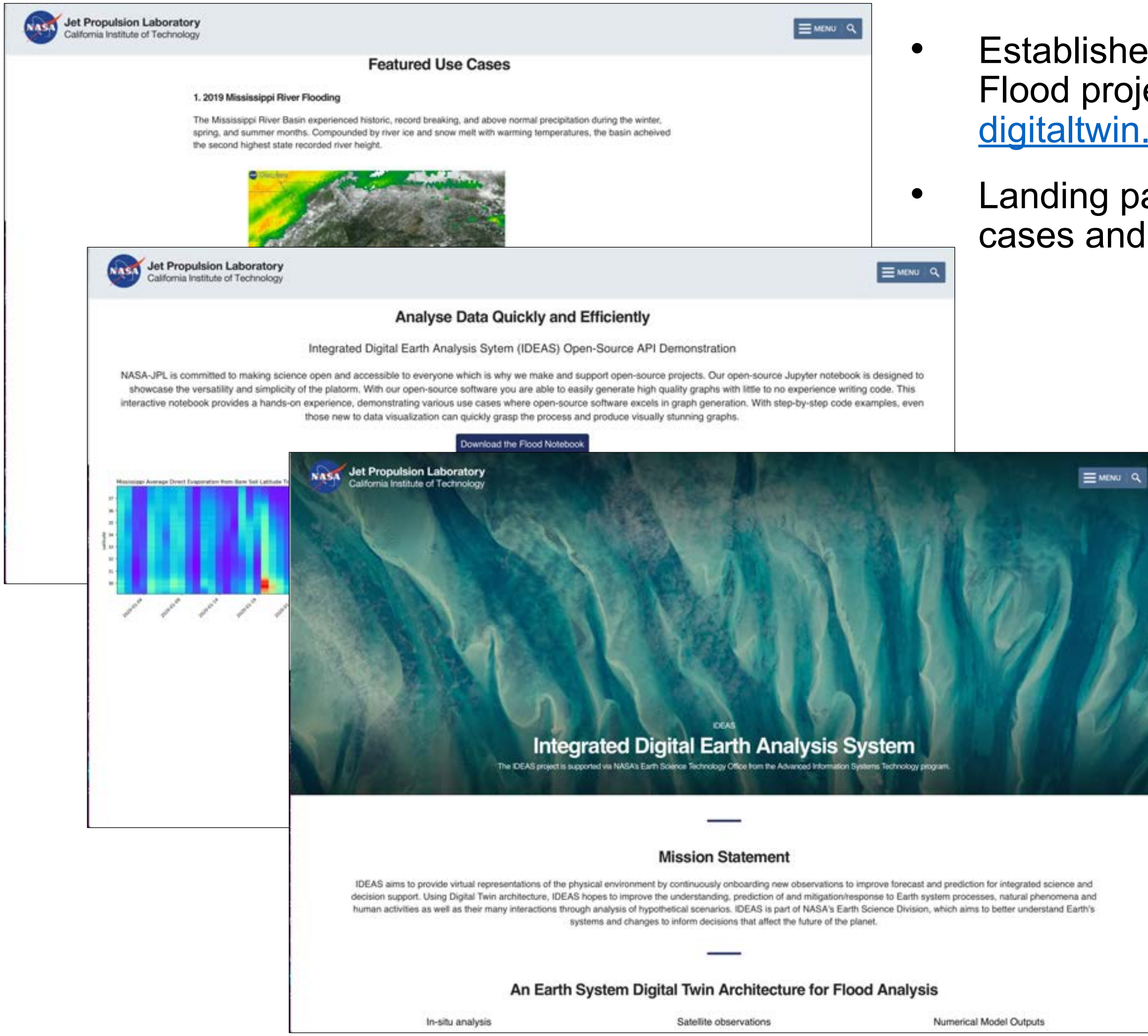
IDEAS – Digital Twin for Water Cycle and Flood Detection and Monitoring

Summary of Accomplishments

- New project website
- New datasets
  - Land Surface Models: NOAH, CLSM, and VIC
  - Satellite data: SMAP L3 Soil Moisture
  - RAPID discharge using different LSM inputs
  - LIS (NoahMP) runs to support 1x 2x 3x precipitation scenarios
  - RAPID discharge outputs for 1x 2x 3x precipitation scenarios
  - Vigicrues in situ observations for Garonne
- What-If examples
- Metadata service
- OGC API – Processes
  - LIS integration
  - RAPID integration
- Telemac2D output to CF-netCDF transformer
- Engagements

Earth Science Technology Office

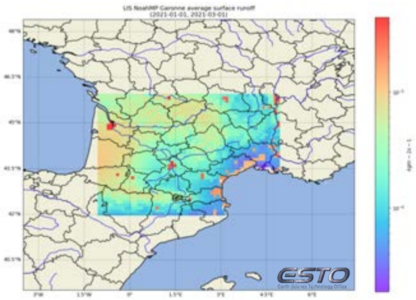
Flood Digital Twins Website



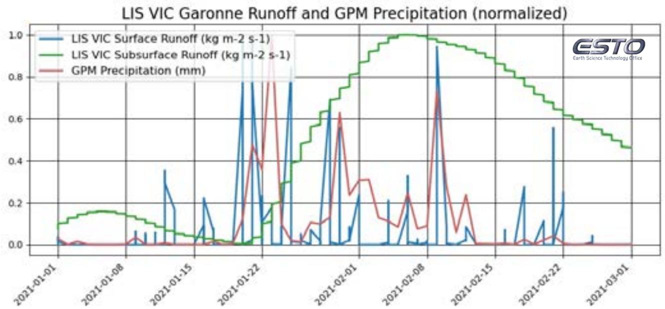
- Established landing page for IDEAS-powered Flood project - http://ideas-digitaltwin.jpl.nasa.gov/floods
- Landing page includes links to notebook, use cases and scenarios

Bringing Observations and Models Together

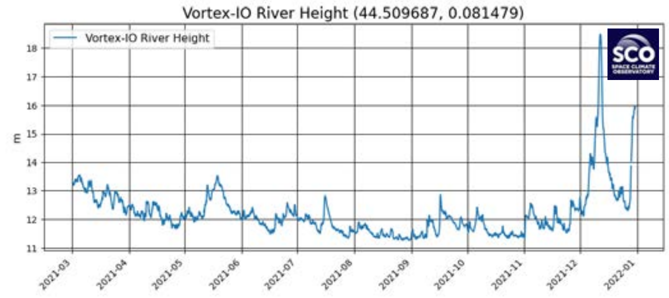
2021-03 through 2021-12 in Garonne



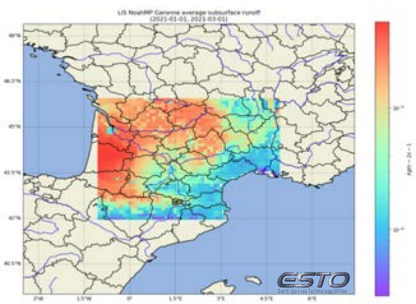
NoahMP Average Surface Runoff



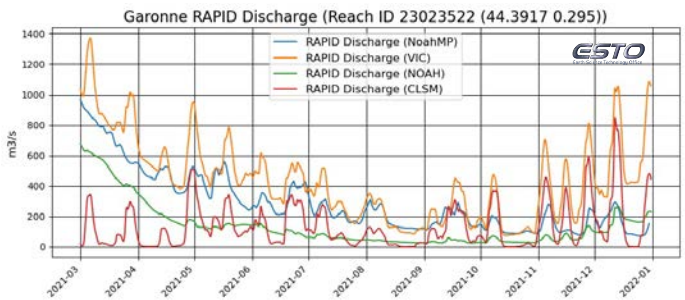
VIC Runoff and GPM Precipitation (normalized)



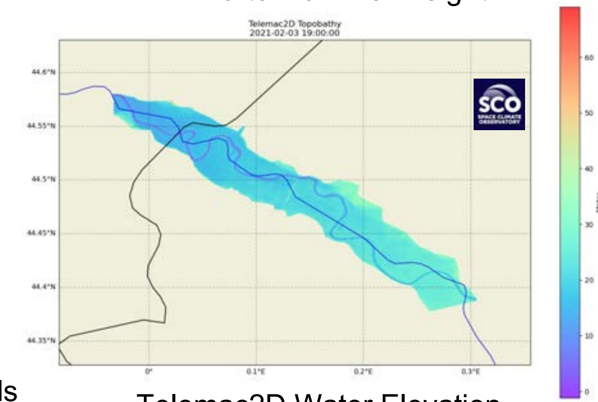
Vortex.Io River Height



NoahMP Average Subsurface Runoff

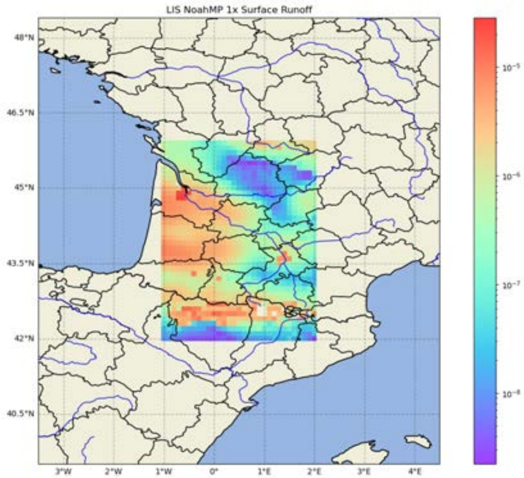


RAPID Discharge from different Land Surface Models

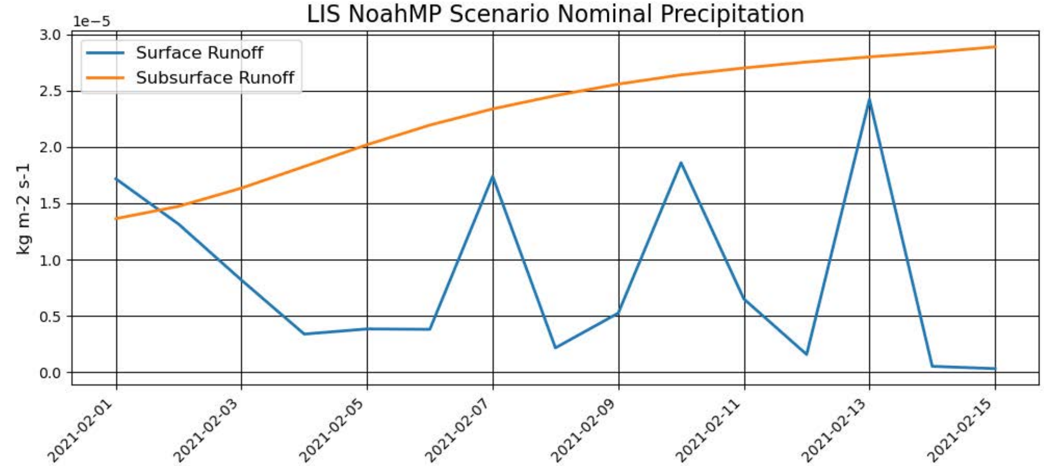


Telemac2D Water Elevation

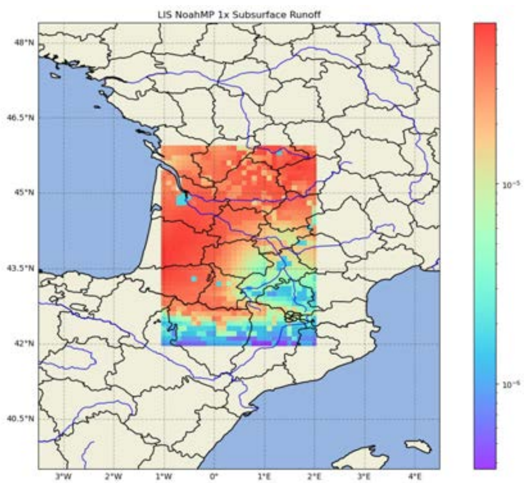
What-If Garonne: Nominal Precipitation



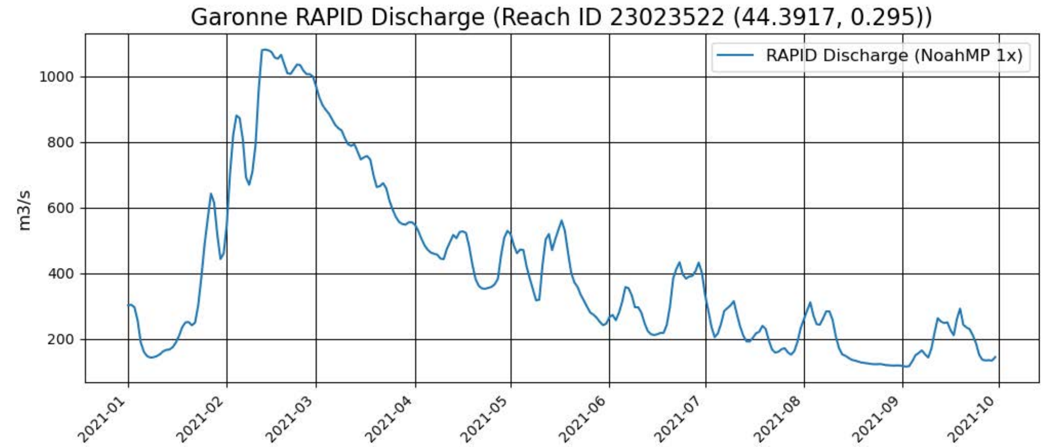
NoahMP – Surface Runoff



NoahMP Surface and Subsurface Runoffs

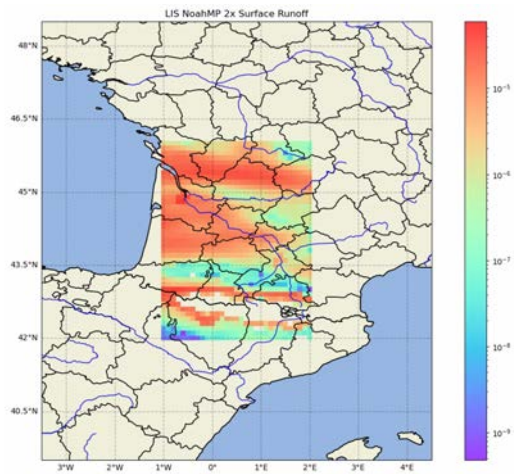


NoahMP – Subsurface Runoff

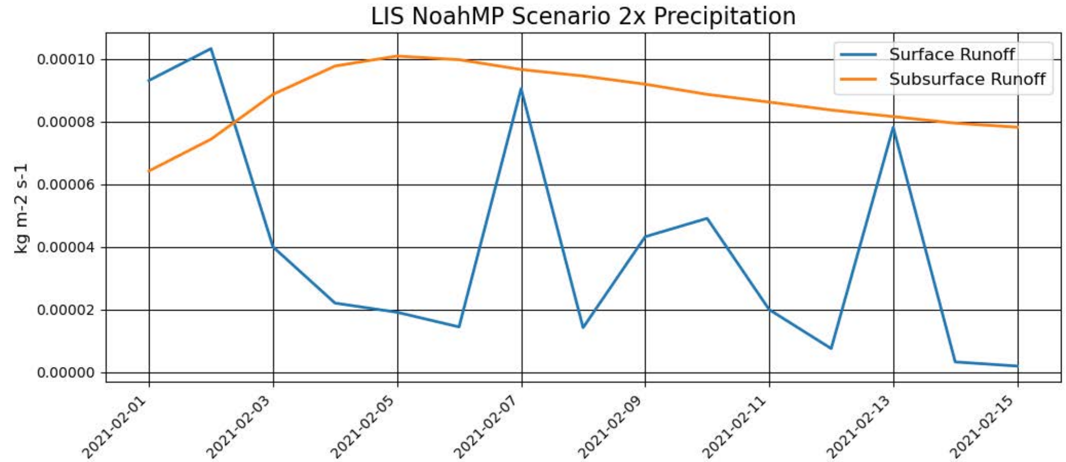


RAPID Discharge

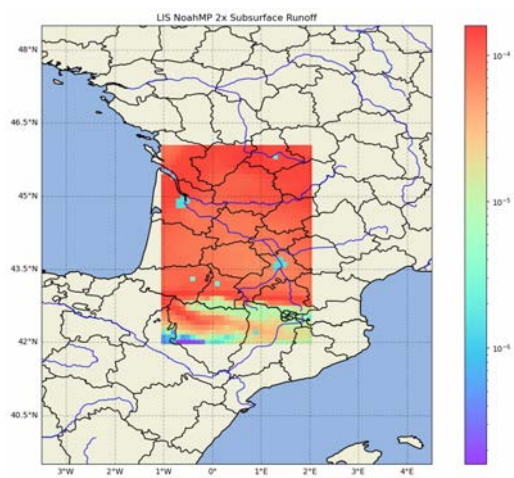
What-If Garonne: 2x Precipitation



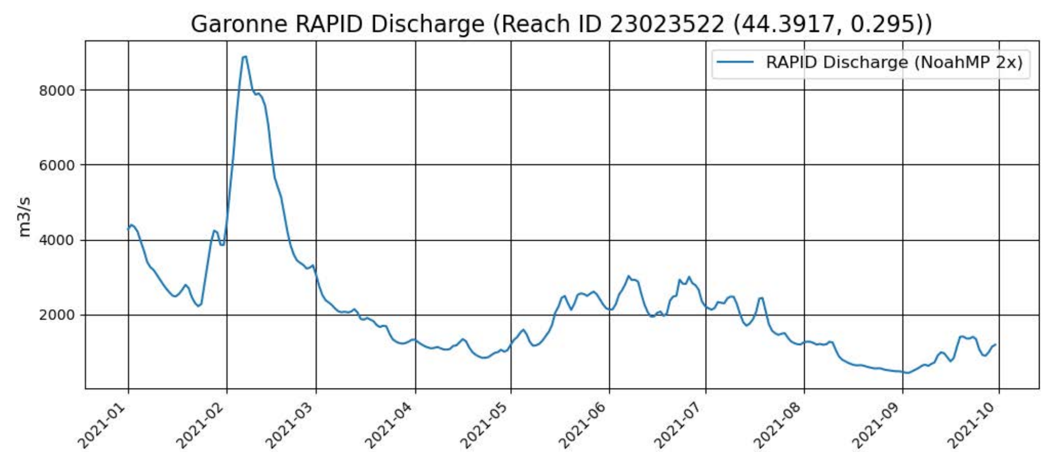
NoahMP – Surface Runoff



NoahMP Surface and Subsurface Runoffs

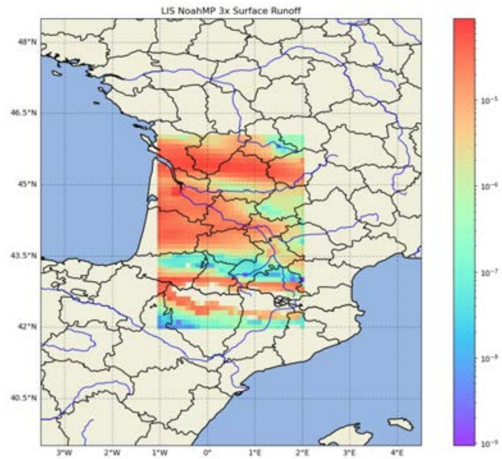


NoahMP – Subsurface Runoff

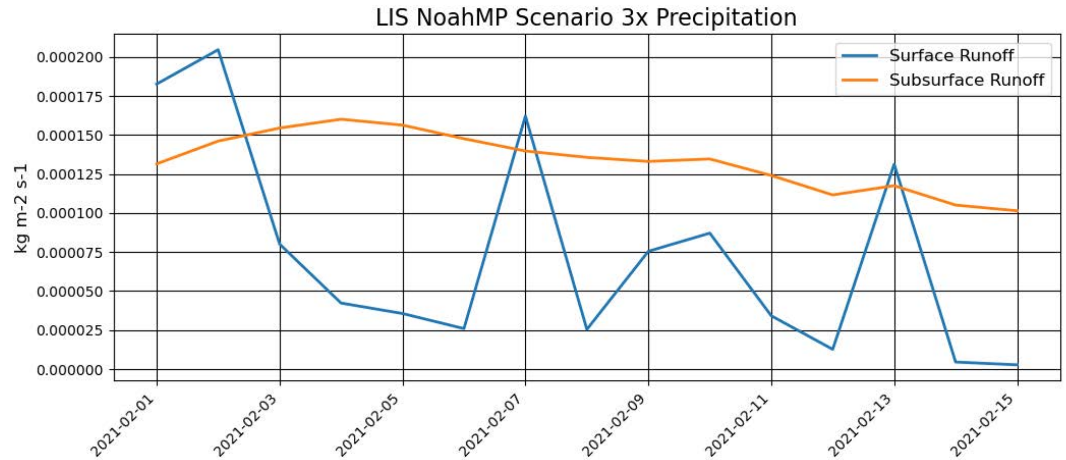


RAPID Discharge

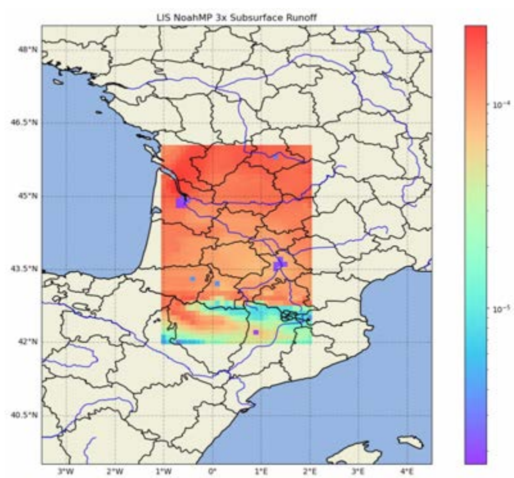
What-If Garonne: 3x Precipitation



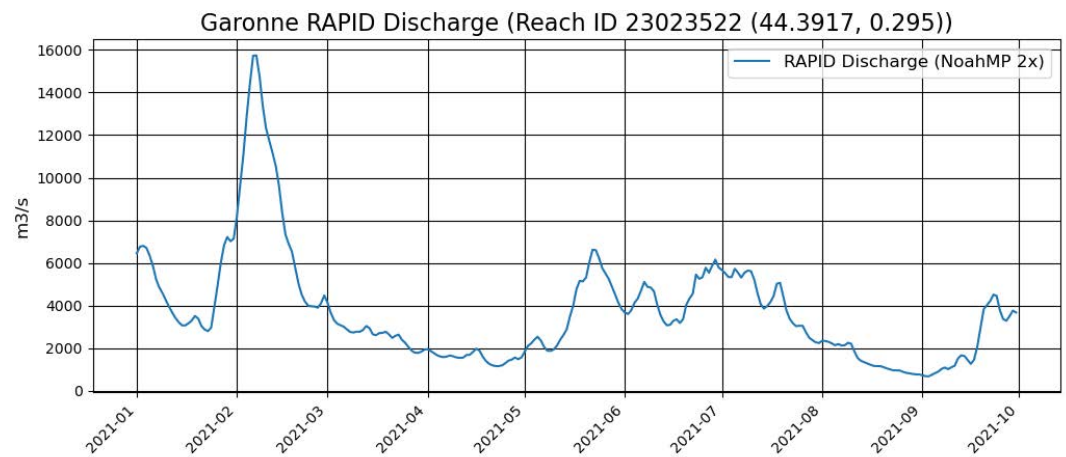
NoahMP – Surface Runoff



NoahMP Surface and Subsurface Runoffs



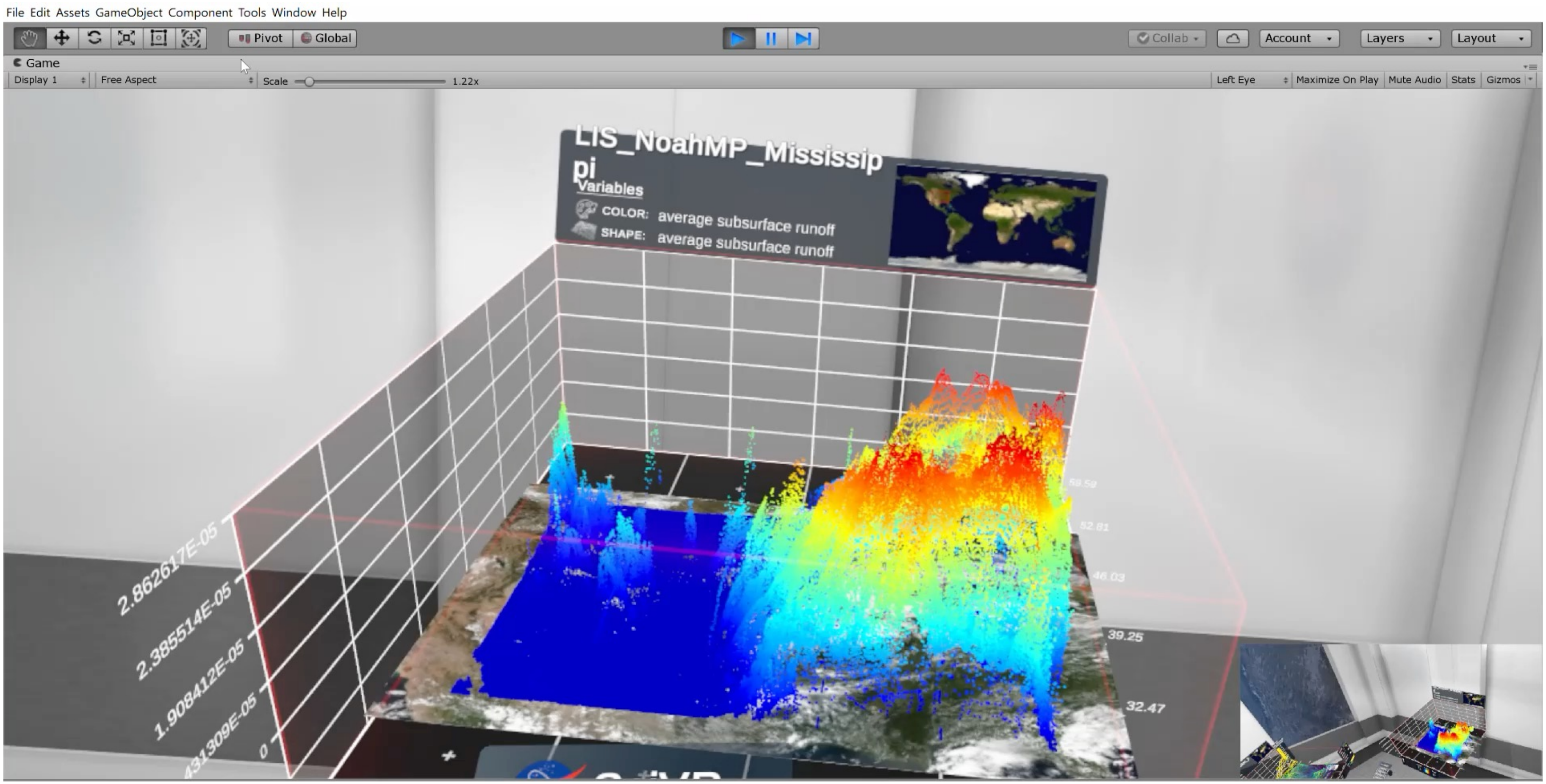
NoahMP – Subsurface Runoff



RAPID Discharge

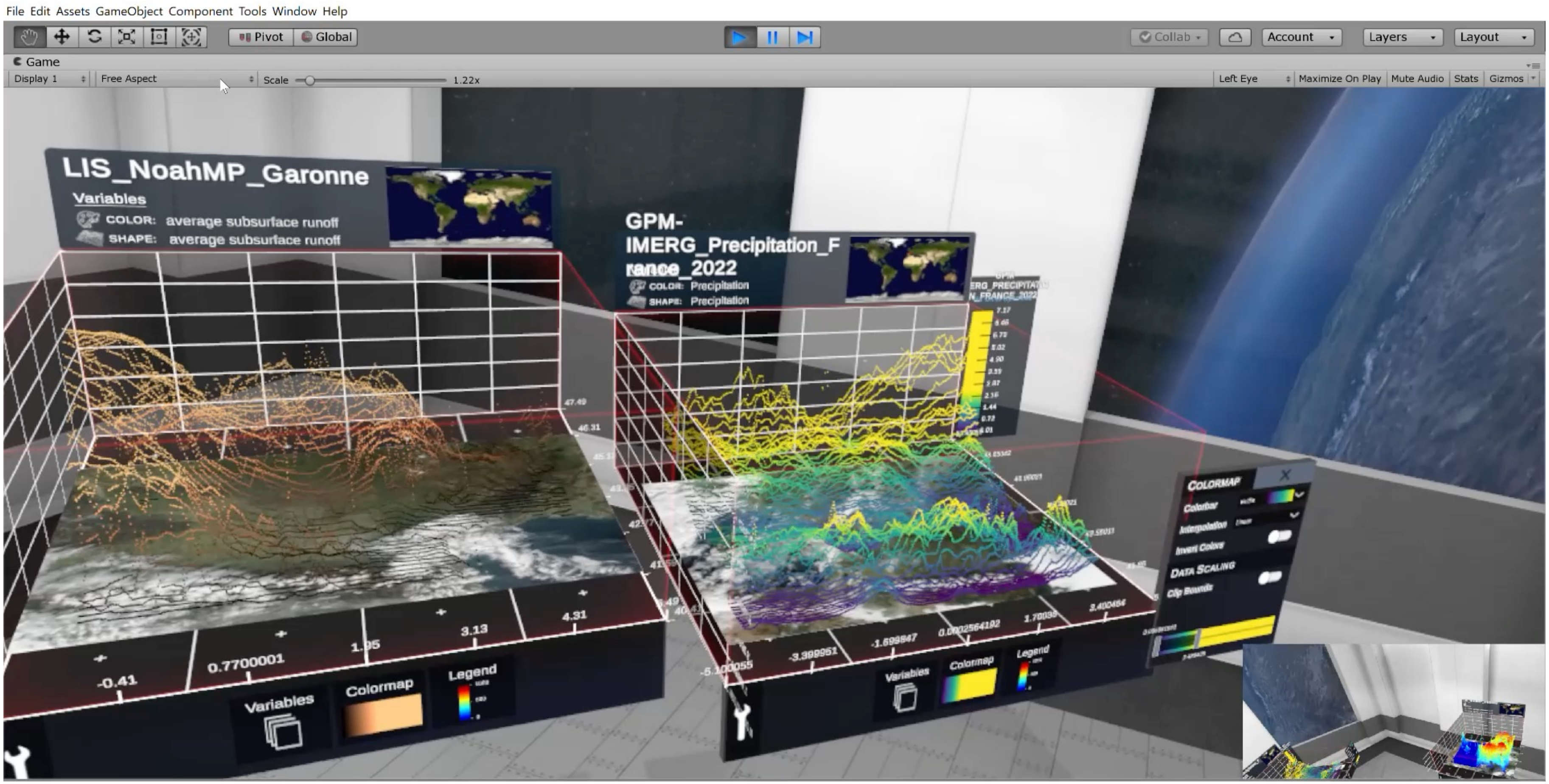
Immersive Science for Flood

Accomplishment



Mississippi NoahMP 2019-02-14 to 2019-08-31

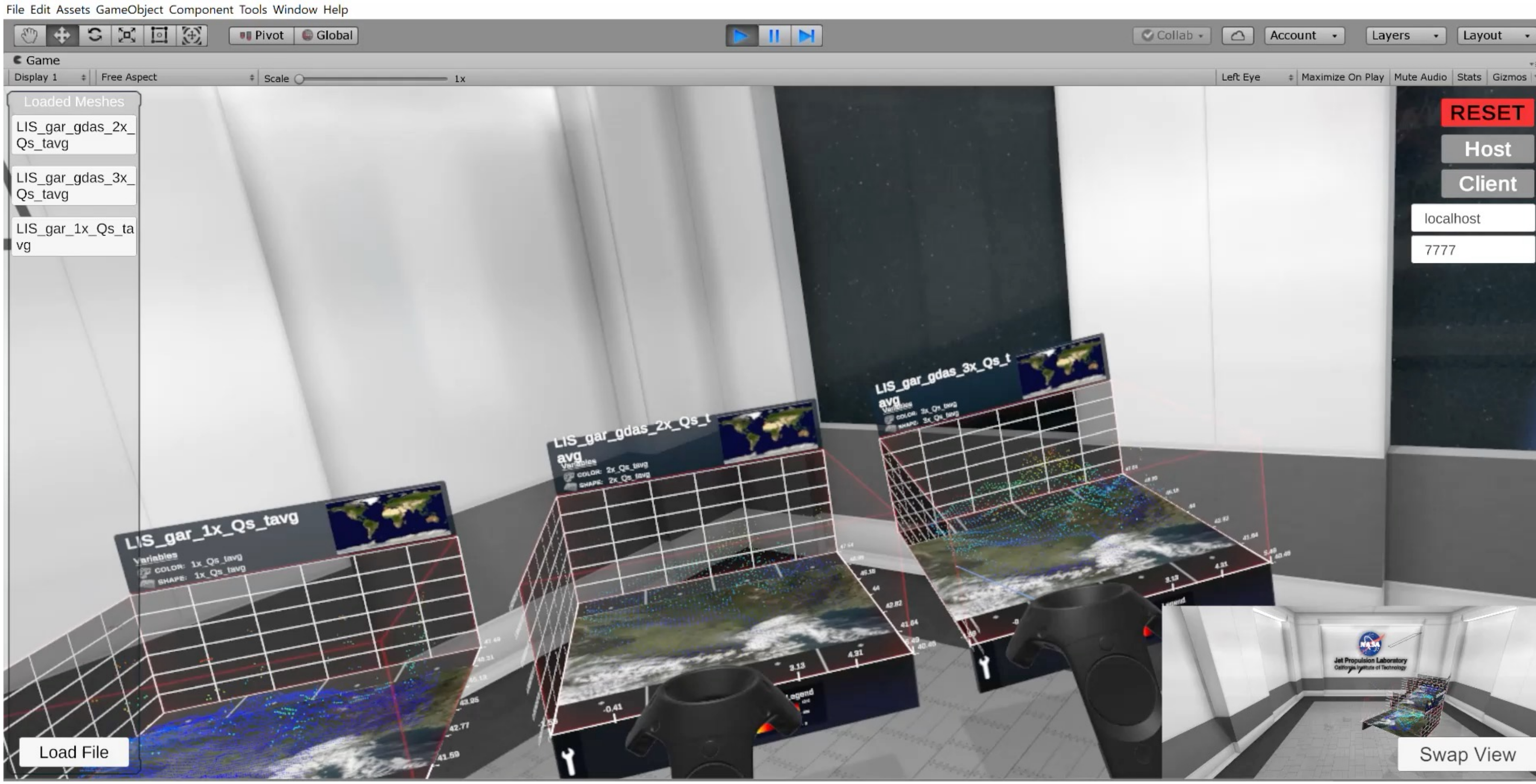
Immersive Science for Flood
Accomplishment



Storm in France 2021-01-01 to 2021-03-01

ESTO Earth Science Technology Office

Immersive Science for Flood
Accomplishment

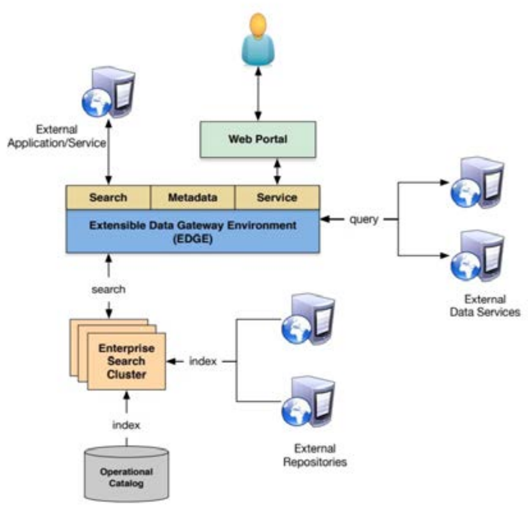


Different NoahMP Scenario Outputs

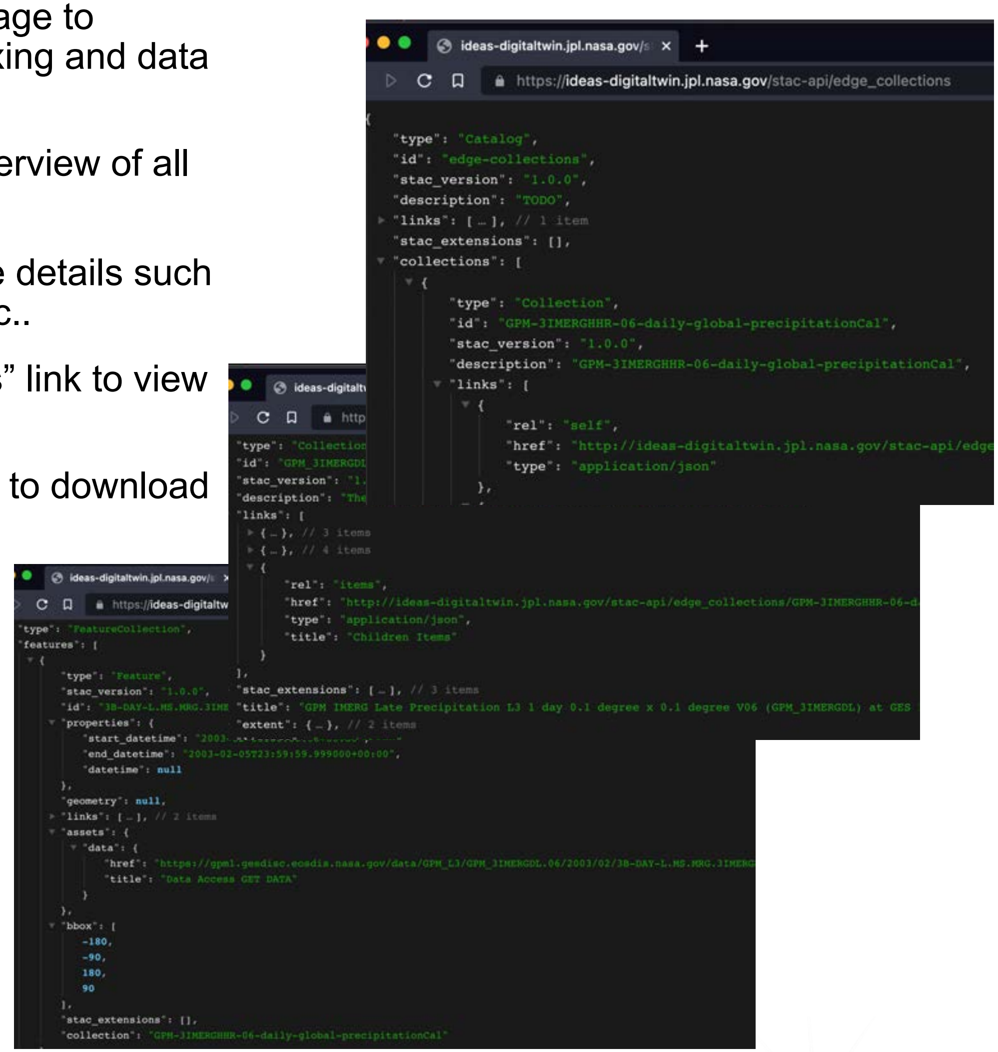
ESTO Earth Science Technology Office

IDEAS Metadata Service and STAC

- STAC – specification for common language to describe geospatial information for indexing and data search
- Base URL at /edge_collections to list overview of all ingested collection.
- Each collection has “self” link to view the details such as description, summary, assets, and etc..
- (Update) Each collection also has ”items” link to view all ingested granules for the collection.
- (Update) Each granule has an asset link to download the data file

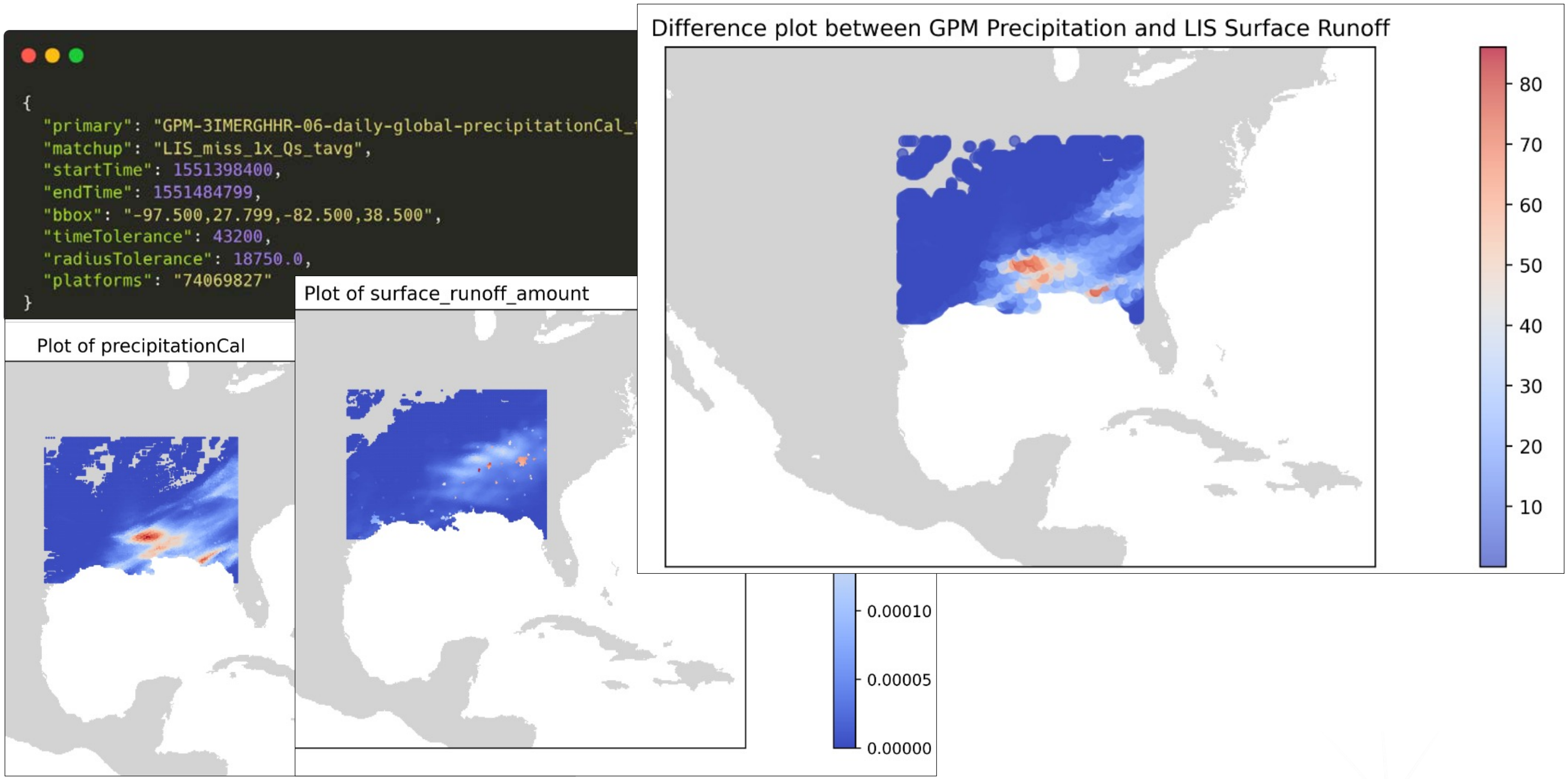


Metadata Architecture



New Match-Up Service

- Collaborated with NASA ACCESS Cloud-Based Data Matchup Service (CDMS) project to integrate matchup capability with Flood data holdings.
- Notebook demo: https://github.com/access-cdms/cdms-notebooks/blob/master/CDMS-AQ-Match-Up-Demo.ipynb



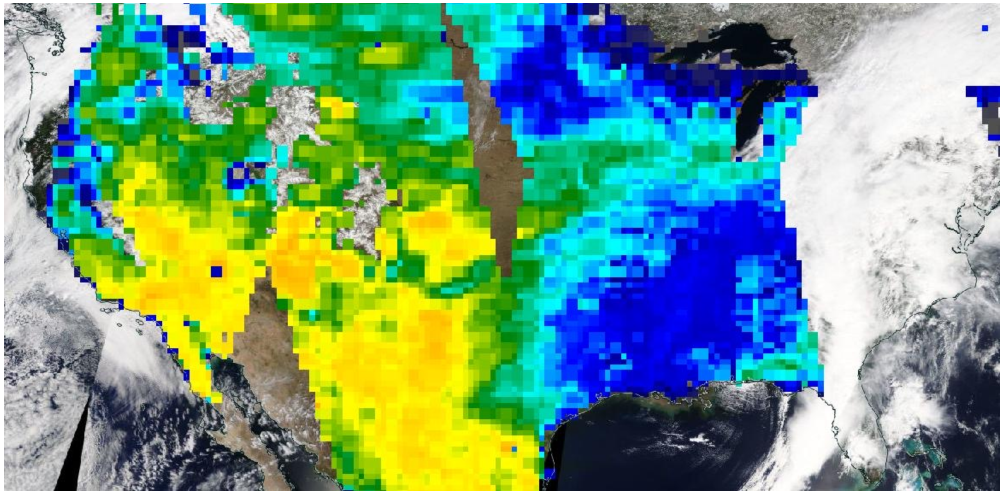
Data Products
Accomplishment

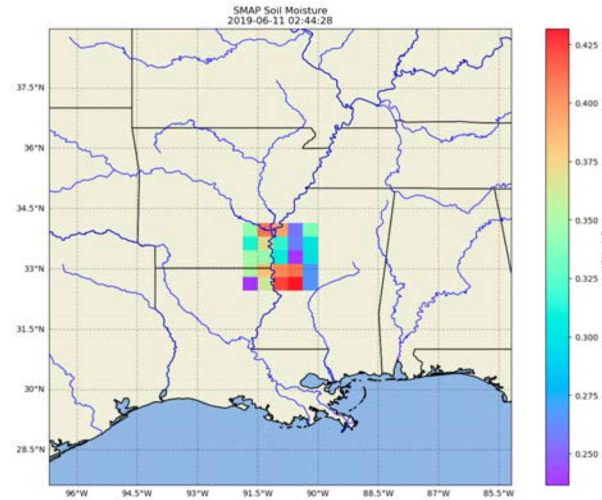
Product	Type	Time Coverage	Location
LIS Surface and Subsurface Runoff Average Daily	Model	2022-05-10 – 2022-05-31	Mississippi River Basin and Garonne River Basin
VIC Surface and Subsurface Runoff Accumulation 3 Hourly	Model	2018-10-01 – 2022-03-01	Global
CLSM Surface and Subsurface Runoff Accumulation 3 Hourly	Model	2018-10-01 – 2022-03-01	Global
NOAH Surface and Subsurface Runoff Accumulation 3 Hourly	Model	2018-10-01 – 2022-03-01	Global
LIS (NoahMP) Surface and Subsurface Runoff Average Daily (1x, 2x, 3x)	Model	2020-09-01 – 2022-12-31	Mississippi and Garonne
RAPID River Discharge Daily (NoahMP input)	Model	2018-10-01 – 2019-08-31 (Mississippi River Basin) and 2021-01-01 – 2021-04-01 (Garonne River)	Mississippi River Basin and Garonne River
RAPID River Discharge 3 Hourly (VIC input)	Model	2018-10-01 – 2019-08-31 (Mississippi River Basin) and 2021-01-01 – 2021-04-01 (Garonne River)	Mississippi River Basin and Garonne River
RAPID River Discharge 3 Hourly (CLSM input)	Model	2018-10-01 – 2019-08-31 (Mississippi River Basin) and 2021-01-01 – 2021-04-01 (Garonne River)	Mississippi River Basin and Garonne River
RAPID River Discharge 3 Hourly (NOAH input)	Model	2018-10-01 – 2019-08-31 (Mississippi River Basin) and 2021-01-01 – 2021-04-01 (Garonne River)	Mississippi River Basin and Garonne River

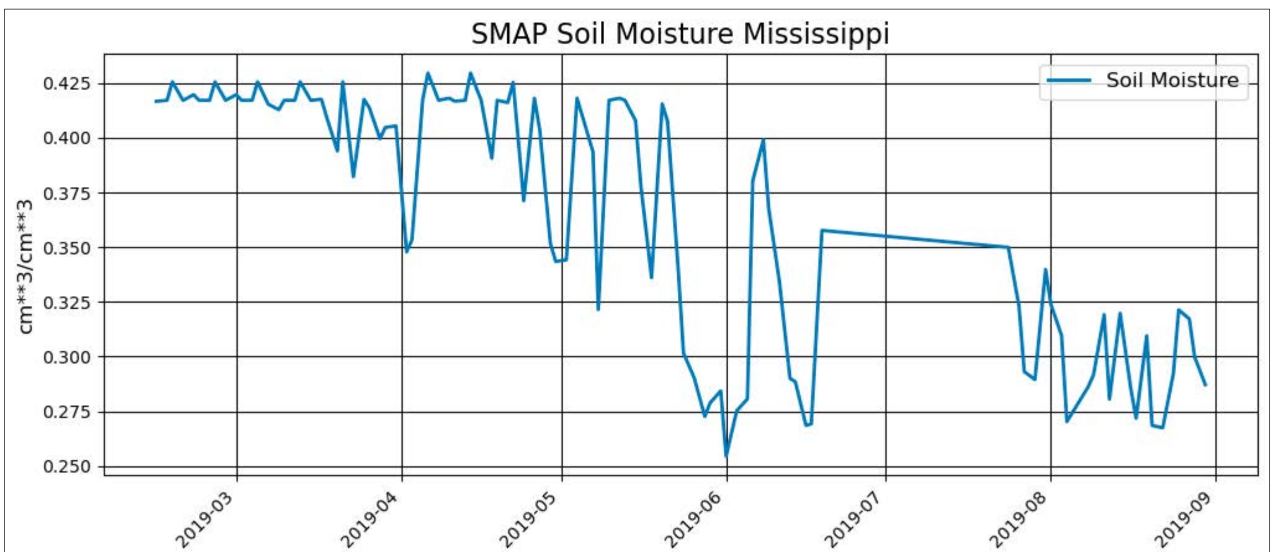
Data Products
Accomplishment

Product	Type	Time Coverage	Location
RAPID River Discharge Daily (NoahMP input) (1x, 2x, 3x)	Model	2020-09-01 – 2022-12-31	Mississippi River Basin and Garonne River
GPM 3IMERGDL_06 Accumulated Precipitation Daily	Satellite	2018-10-01 - 2022-10-02	Global
SMAP SPL3SMP Soil Moisture	Satellite	2018-10-01 – 2019-08-31 (Mississippi River Basin) and 2021-01-01 – 2021-04-01 (Garonne River)	Mississippi River Bain and Garonne River
USGS Water data	In situ	2018-11 – 2019-09	Mississippi River Basin
VorteX.io Micro-Station	In situ	2022-01-01 – 2022-01-29	Garonne River Basin
Vigicrues	In situ	2020-11-01 - 2022-01-29	Garonne River Basin

New SMAP L3 Soil Moisture



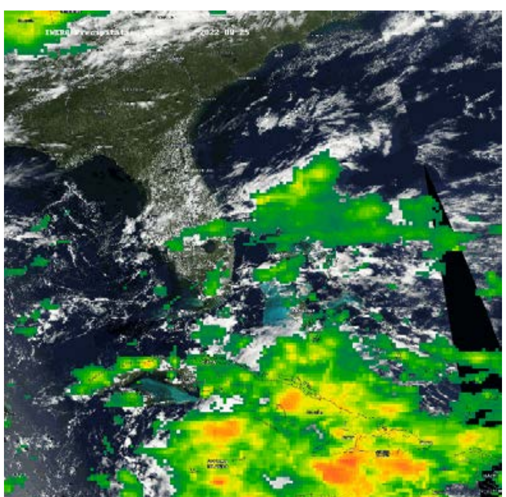


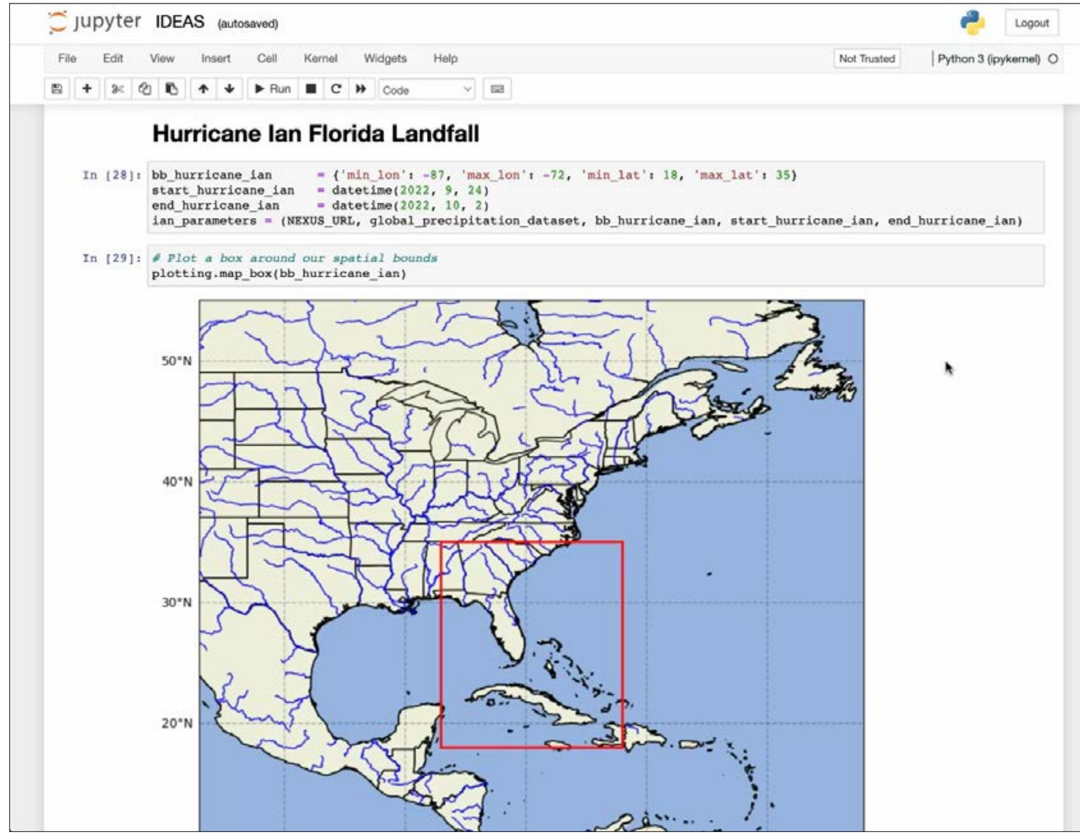


IDEAS-powered Flood Notebook

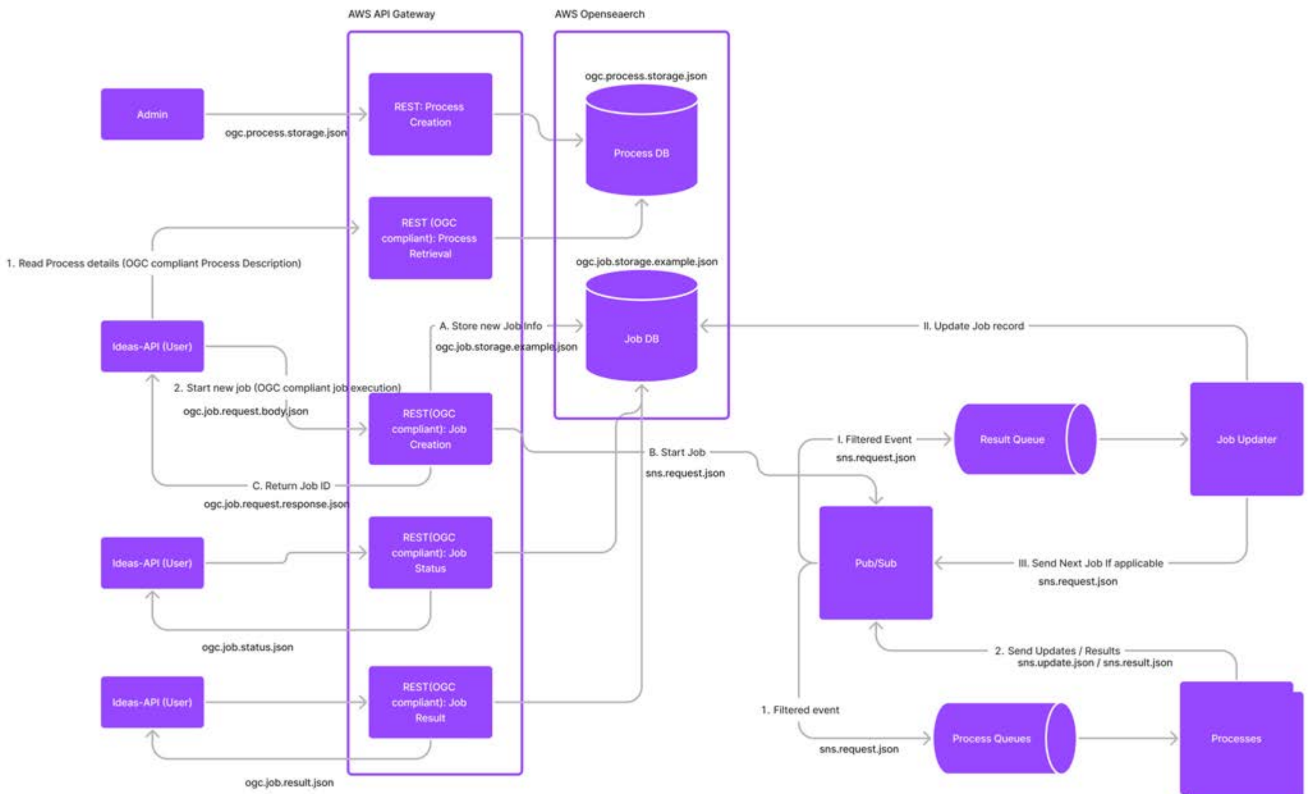
Accomplishment

- New updates to the Flood Notebook to demonstrate access, analysis, and visualization - https://github.com/EarthDigitalTwin/IDEAS-notebooks
- The latest release includes examples for Mississippi, Garonne, and other regions using NoahMP, CLSM, NOAH, VIC, RAPID, USGS, VorteX, and Vigicrues data
- Example accessing precipitations scenarios for Mississippi and Garonne





Data Integration Architecture
Accomplishment

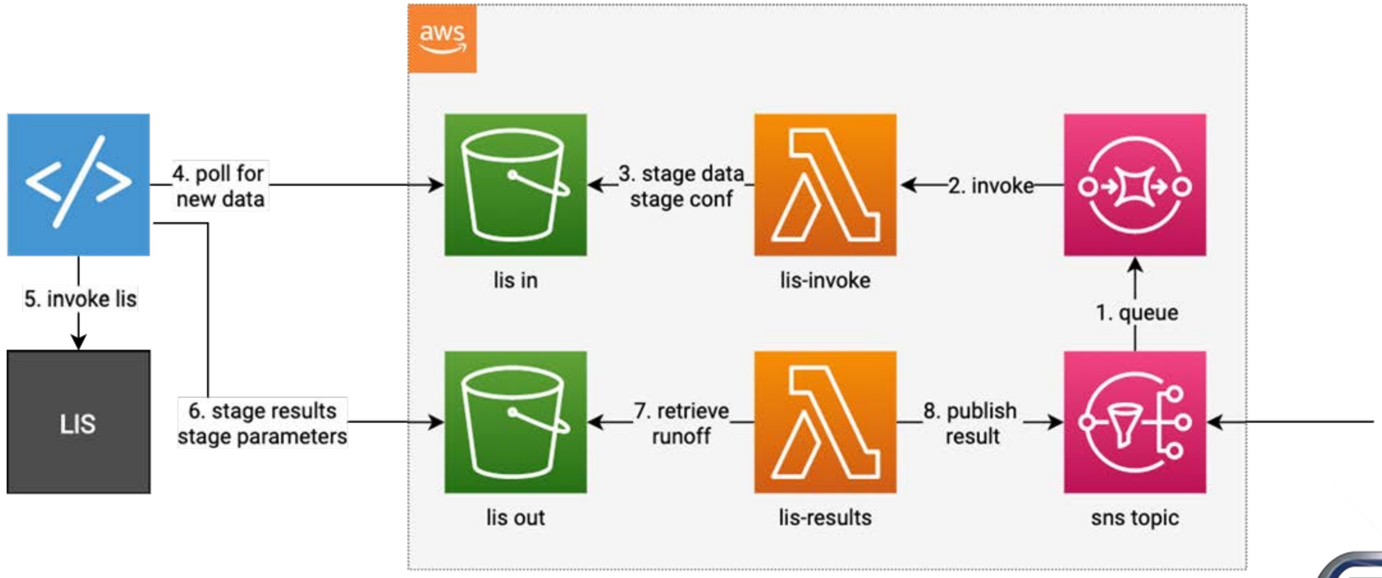


- Focuses
  - Data harmonization
  - Data provenance
  - Seamless integration of data and services
- IDEAS API is compliant with the OGC Processes standard

LIS Integration

Accomplishment

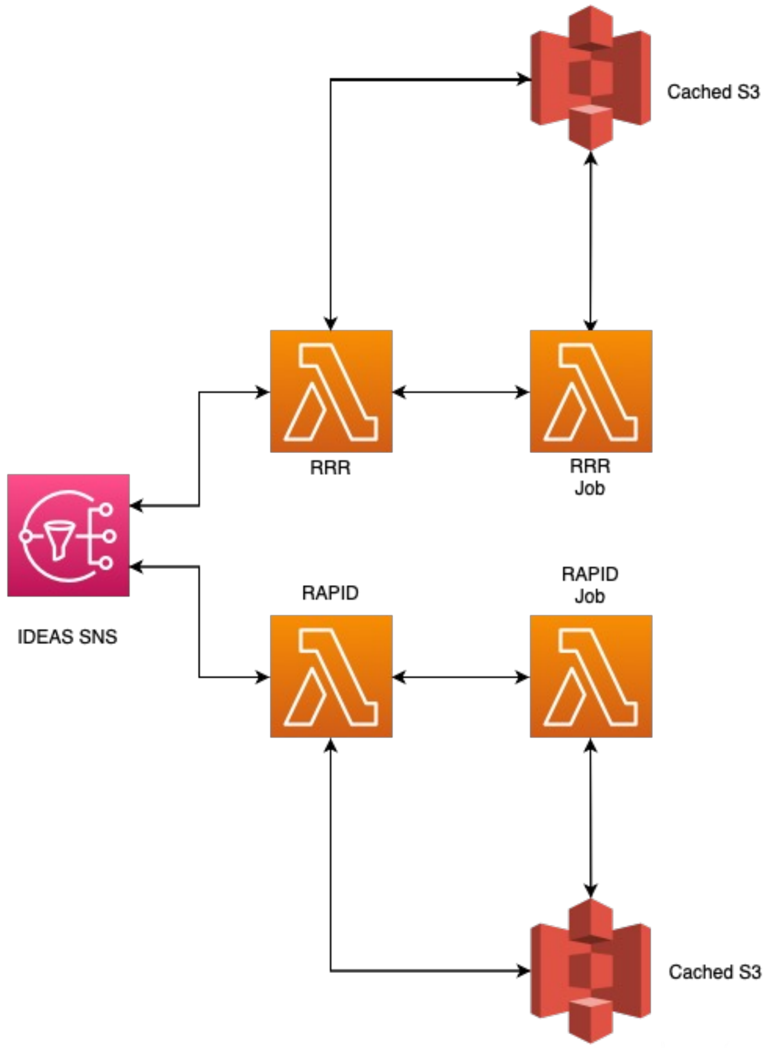
- **Goal:** establish general webservice API to encapsulate the complexities of large-scale model configuration and the execution environment
- **LIS Input**: Precipitation and metrological products
- **LIS Output**: Surface and subsurface runoff product
- LIS is currently hosted on NASA NCCS HPC
- Information model between IDEAS and models includes job lifecycle and data provenance
- Use event-based method for managed-integration between subsystems
  - LIS is invoked by a cloud-native event-driven system
  - Parameters used to run LIS will be maintained for provenance
  - After completion, publication to SNS topic will trigger next subsystem (e.g., RAPID)
- Input is in OGC Processes format. The LIS lambda converts the OGC format to a simplified format which the LIS cron converts to an LIS HPC job.



RAPID Cloud Deployment and Integration

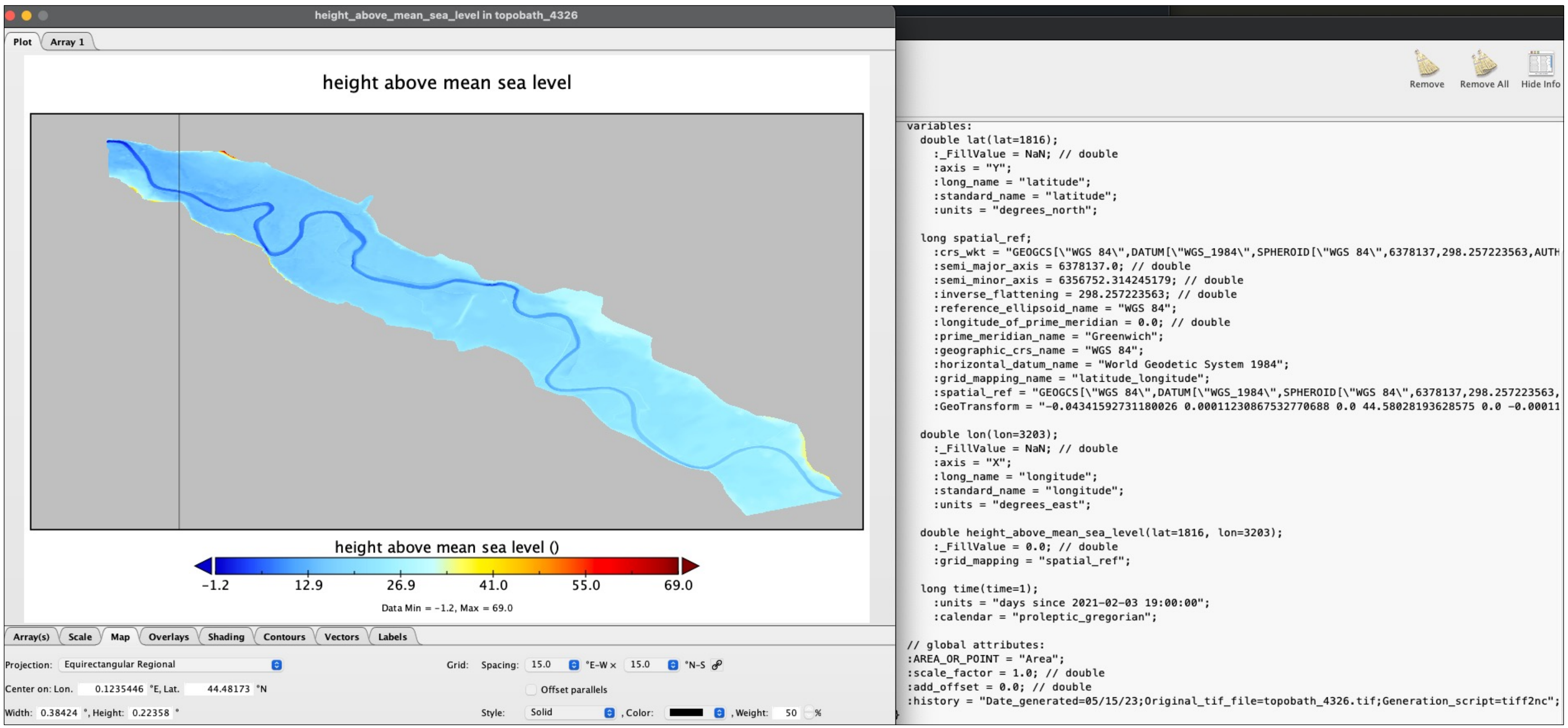
Accomplishment

- Containerized RRR, RAPID's ETL pipeline, for data transformation and preprocessing
- Containerized RAPID model for on-demand execution
- Leverage serverless solution for webservice API (AWS Lambda)
- Both RRR and RAPID results are caches in AWS S3
- Orchestration between RRR and RAPID managed by IDEAS OGC Processes API



Telemac2D GeoTiff to CF-netCDF Transformer

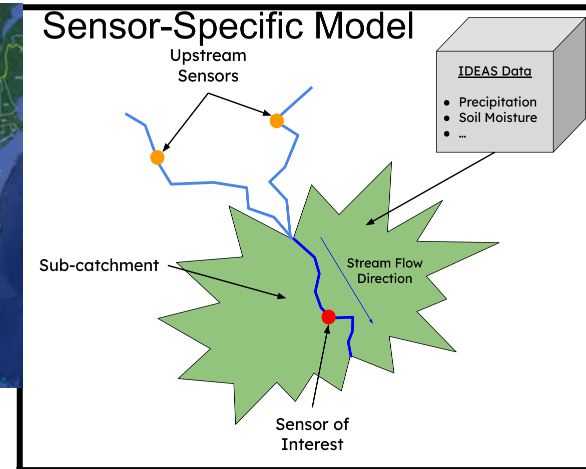
- Tool developed to converted Telemac2D geotiff data to CF Compliant NetCDF files

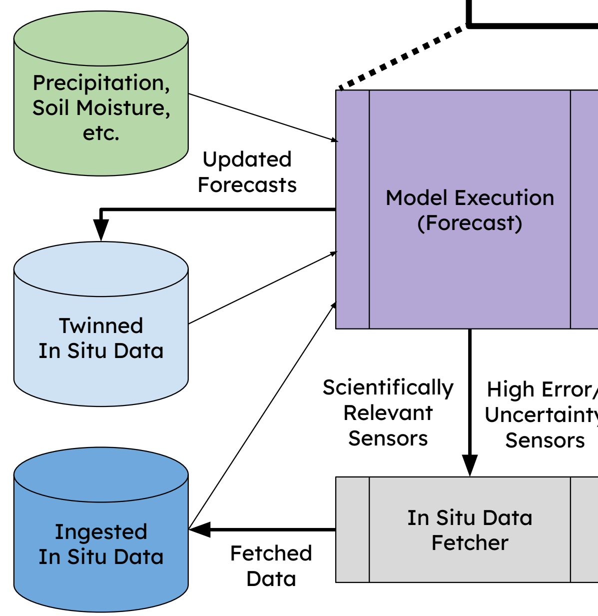


ESTO Earth Science Technology Office

ML-Driven In-Situ Data Acquisition

- **Objective:** selectively fetch in situ sensor data corresponding to events of interest (e.g., flooding)
- **General Approach:**
  - Use twin to forecast sensors with expected high stream gauge readings
  - Use forecast to inform fetching of new in situ measurements
  - Update internal state with new measurements and repeat process at next forecast period (~6-24 hours)
- **Modeling Approach:**
  - Long short-term memory (LSTM) neural network models estimate stream gauge readings given upstream readings and precipitation (IMERG GPM)
  - GPM data mapped from pixels to MERIT basins
  - Historical data used for training





Next Steps:

- Integrate model training and execution into AWS
- Develop continuous re-training
- Add soil moisture and other datasets

Machine Learning Modeling Details

Dataset Curation:

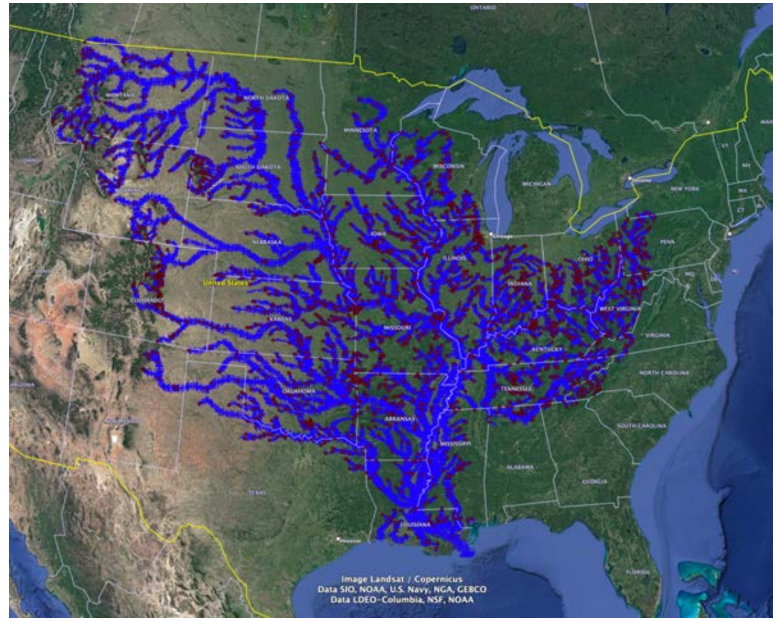
- Matched USGS in situ sensors (~2500) to MERIT river reach and basin database
- Build network of upstream sensors/basins for each sensor
- Generate mask to extract basin-specific GPM data

Input Training Data:

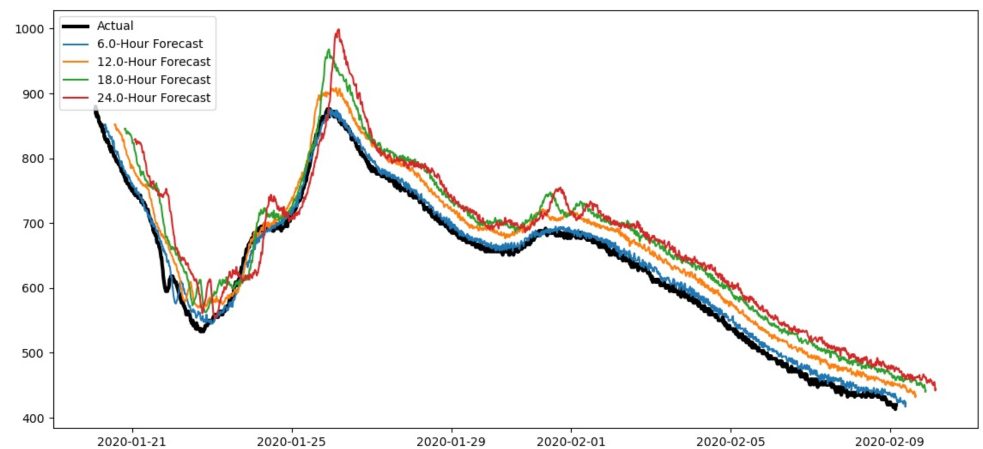
- 15-minute historical USGS readings
- Interpolated daily GPM data

Forecast Model:

- LSTM Timeseries Regression Model
- Ongoing: characterizing trade-off between performance and prediction interval



MERIT River Network and matched USGS sensor locations



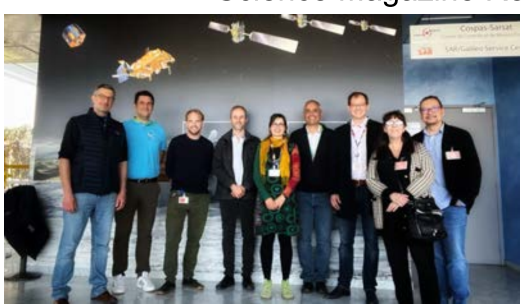
ESTO Earth Science Technology Office

New Collaboration

- Colin Gleason (UMass Amherst) on SWOT discharge product available at the PO.DAAC (Oct 2023)
  - Ref: Suresh Vannan (Luisa Vieira Lucchese) *A Hosted Analytic Collaborative Framework for Global River Water Quantity and Quality from SWOT, Landsat, and Sentinel-2*

Recent Engagements and Highlights

- CEOS 55th Meeting of Working Group on Information Systems and Services (WGISS), Cordoba, Argentina
  - Invited to join the CEOS WGISS
  - Invited to use IDEAS to formalize the CEOS WGISS Interoperability Framework
- CNES Toulouse Meeting
- EGU 2023, Vienna, Austria
- Group on Earth Observations (GEO) Dialogue Series on Open Software, Open Infrastructure, and Open Hardware
- EU's Standards in Digital Twins of the Ocean
- Upcoming
  - Oct. 19 and 20, 2023, HAQAST Public Meeting in Salt Lake City, Utah
  - IGARSS 2023 Hyperwall
  - IGARSS 2023 – "The SCO-FloodDAM Project: Towards a Digital Twins for Flood Detection, Prediction and Flood Risk Assessments"
  - IGARSS 2023 – "Big Data Smart: Federated Earth System Digital Twins"
  - ESIP Summer Meeting
  - Bulletin of the American Meteorological Society
  - Science Magazine Review Article



ESTO Earth Science Technology Office

Q&A

Clearance: CL#23-3227

NASA AIST Fire Alarm
Science Data Platform for Wildfire and Air Quality
Mid-year Review

Thomas Huang[1], Gary Doran[1], Sina Hasheminassab[1], Sarah Hallam[1] (Intern), Olga Kalashnikova[1], Jason Kang[1], Kyo Lee[1], Grace Llewellyn[1], Kevin Marlis[1], Jessica Neu[1],

Joe Roberts[1] (Dev Lead), Jeanne Holm[2], Mohammad Pourhomayoun[3], Dawn Comer[2]

Chaowei Yang[4], Qian Liu[4], Hai Lan[4], Anusha Srirenganathanmalarvizhi[4]

[1] NASA Jet Propulsion Laboratory
[2] City of Los Angeles
[3] California State University, Los Angeles
[4] George Mason University

Jet Propulsion Laboratory
California Institute of Technology

GEORGE MASON UNIVERSITY

ESTO Earth Science Technology Office

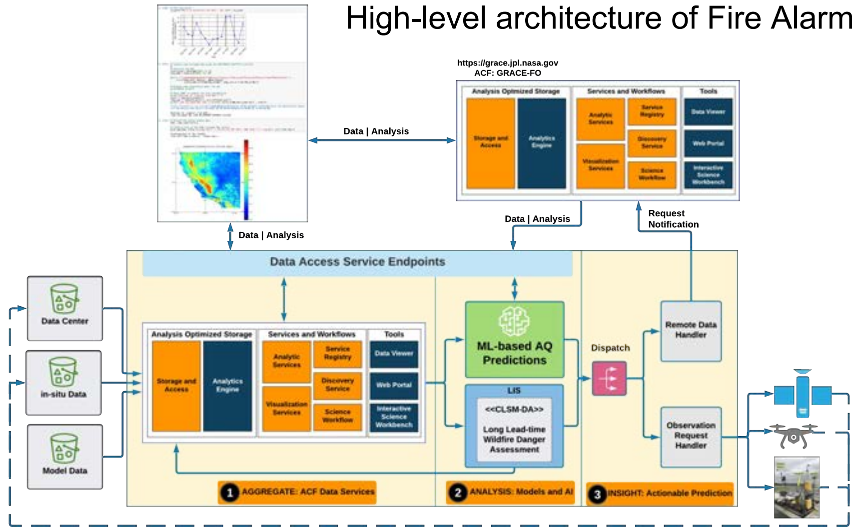
Fire Alarm: Science Data Platform for Wildfire and Air Quality

PI: Thomas Huang, JPL

Objective

- Establish an Analytics Collaborative Framework (ACF) to study the cause and effect of wildfire in relation to air quality and health impacts
- Improve data-driven predictions on wildfire-caused air pollution and health impact
- Align with future SBG fire detection and fire radiative power data for fire risk, detection, and analysis
- Align with future air quality missions such as MAIA
- Solidify ACF-driven NOS-T architecture for dynamic data acquisition and instrument re-tasking
- Integrate and extend the Integrated Digital Earth Analysis System (IDEAS), ESDT framework, for science-driven, scenario-based analysis

High-level architecture of Fire Alarm



Approach

- Expand the AIST Fire Alarm prototype to support relevant fire and air quality measurements for wildfire, air quality, and health impact analysis
- Establish automated workflow solution for near real-time onboarding and transformation of the latest measurements
- Expand the Ocean Carbon Cycle NOS solution to support dynamic data acquisition and re-tasking
- Improve the AIST Fire Alarm prototype's interfaces to enable predictive models of air quality and health impact to be driven by harmonized data from multiple instruments
- Expand the Apache SDAP multiband Analysis Optimized Storage (AOS) capability for hyperspectral data

Co-Is/Partners: Nga T. Chung/JPL, Jessica Neu/JPL, David Schimel/JPL, Olga Kalashnikova/JPL, Mohammad Pourhomayoun/CSULA, JT Reager/JPL, David Diner/JPL, Jeanne Holm/City of LA

Key Milestones

Analysis Optimized Storage (AOS) for hyperspectral measurements and ground-level hotspots matchup	May-23
Scenario-based air quality prediction	Aug-23
Analysis and ML-driven data acquisition (NOS)	Oct-23
Visualization to support scenario-based wildfire-related air quality and health impact	Nov-23
Prototype MAIA retasking	Nov-23

TRL_{in} = 4

Earth Science Technology Office

Summary of Accomplishments

- Official project website: https://ideas-digitaltwin.jpl.nasa.gov/aqacf/
- Air Quality Analysis Tool: https://ideas-digitaltwin.jpl.nasa.gov/aqacf/dat/
- Air Quality Jupyter Notebook: https://github.com/EarthDigitalTwin/FireAlarm-notebooks/
- Virtual Reality support
- TROPOMI global product generation
- HQ demonstration
- Engagements
  - ESIP Winter Meeting
  - CEOS 55th Meeting of the Working Group on Information Systems and Services (WGISS)
  - CNES Toulouse meeting
  - EGU 2023
  - EU's Iliad webinar: Standards in Digital Twins of the Ocean
  - GEO Dialogue Series
- Upcoming Engagements
  - IGARSS
  - ESIP Summer Meeting
  - AMS Interview
  - Science Magazine Review article
  - AGU digital twins session

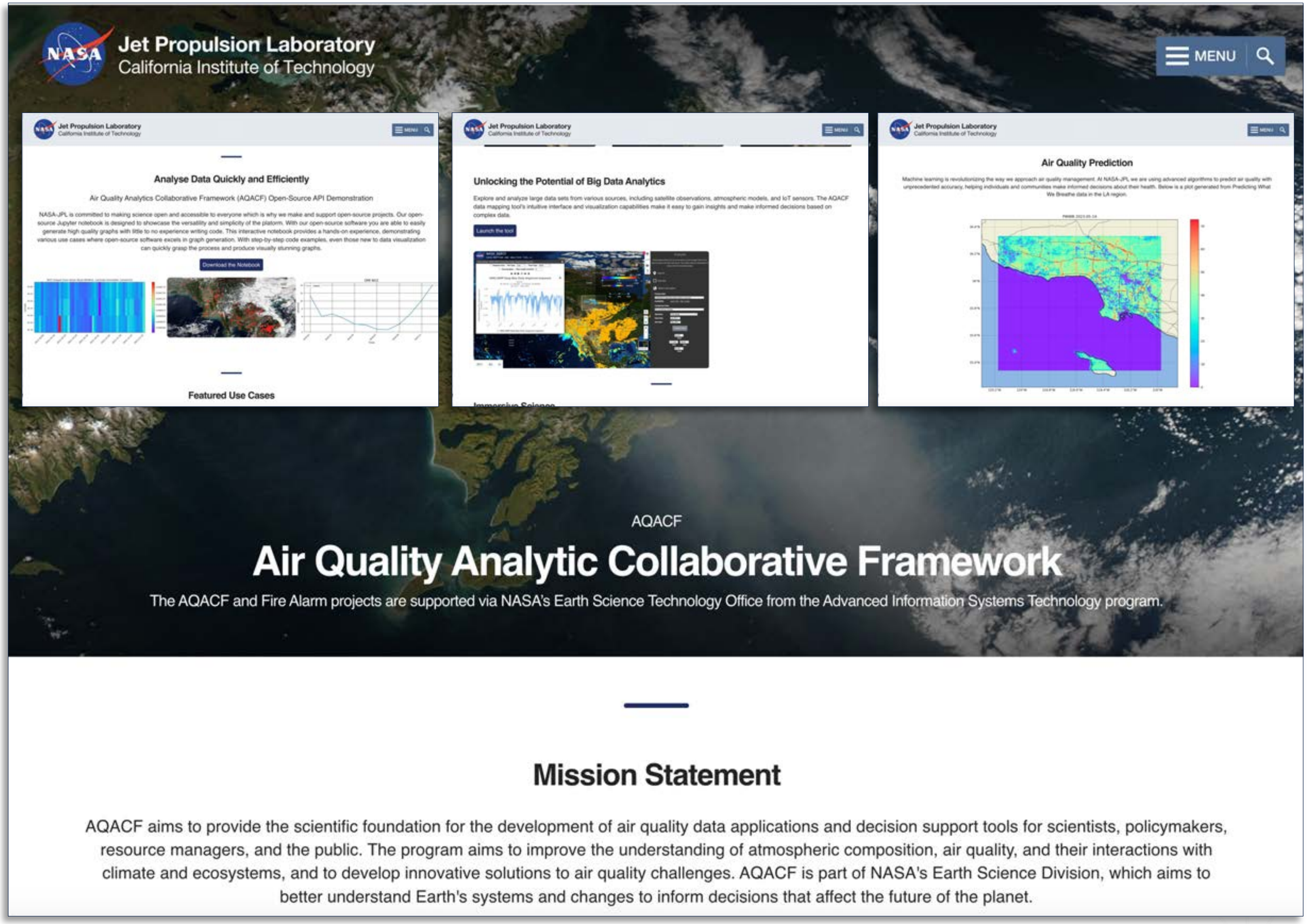
Fire Alarm Development Plan

23.1	23.2	23.3	23.4
Analysis Optimized Storage • In-situ data support	**Analysis Optimized Storage** • In-situ schema updates	**Analysis Optimized Storage** • Hyperspectral data support • Zarr support	**MAIA Retasking** • Retask MAIA observations
Visualization and Analysis • Common metadata services • Jupyter notebook examples	**Visualization and Analysis** • z-dimension support • VR Immersive Science • Wildfire analysis	**Visualization and Analysis** • Greenhouse gas analysis • Scenario-based visualization	**Visualization and Analysis** • Release open source tools
Machine Learning Prediction • Integrate PM2.5, O3, NO, NO2, CO and other air pollutants	**Machine Learning Prediction** • Increase ML spatial resolution • Expanded coverage predictions	**Machine Learning Prediction** • Wildfire air quality predictions • Production air quality predictions	**Machine Learning Prediction** • ML Jupyter notebook examples
Scenario-Based Prediction • Open source health impact calculator	**Scenario-Based Prediction** • Deploy health impact calculator	**Scenario-Based Prediction** • ML driven prediction of health impacts	**New Observing Strategy** • Analysis Driven Data Integration
Data Products • Global TROPOMI • VIIRS/MODIS AOD	**Data Products** • Purple Air • AirNow • Aeronet	**Data Products** • GEOS-CF • MISR winds/plumes • GCHP simulations	**Data Products** • Sample SBG

*-

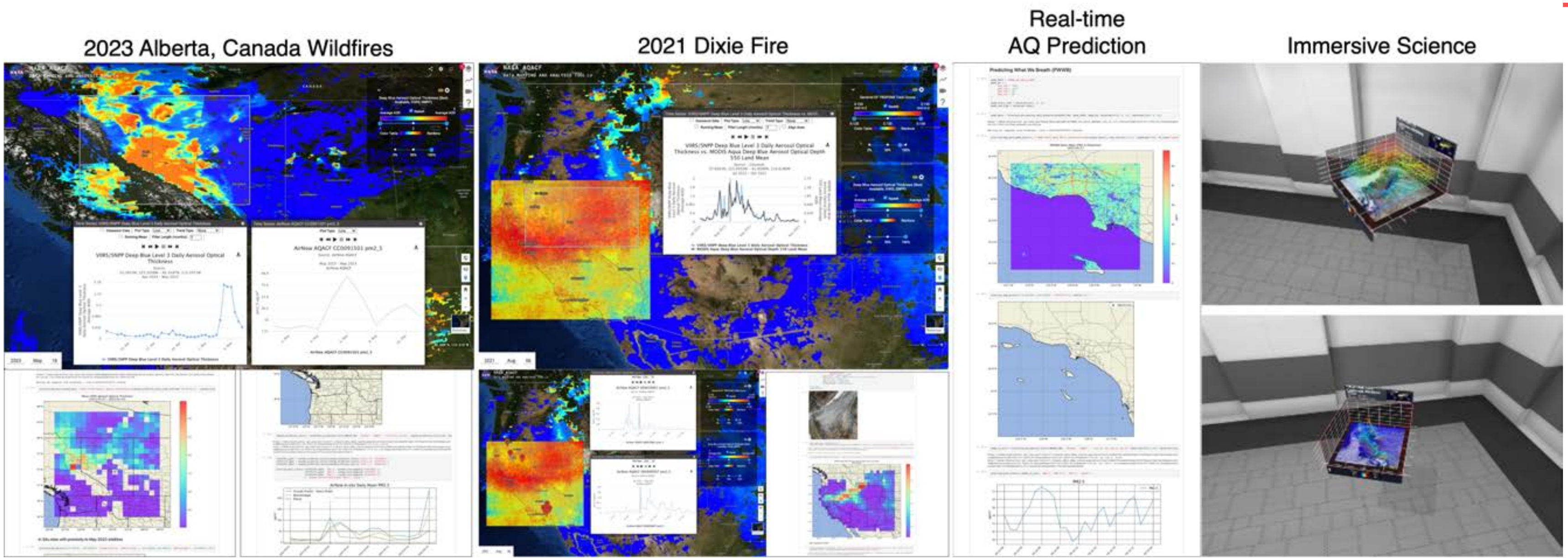
Public-facing Project Website

https://ideas-digitaltwin.jpl.nasa.gov/aqacf/



Earth Science Technology Office

HQ Demonstration



- Provide HQ AQ Program Managers an overview of the professional open-source AQ ACF platform to highlight
  - Technology maturity
  - Ready for distributed analysis – Federated
  - Harmonized data access and analysis
  - Streamline ML training and deployment
- Demonstrations include
  - Web-based data visualization analysis tool
  - Jupyter notebook for AQ
  - PM2.5 daily predictions

ESTO Earth Science Technology Office

Gridded Data Products

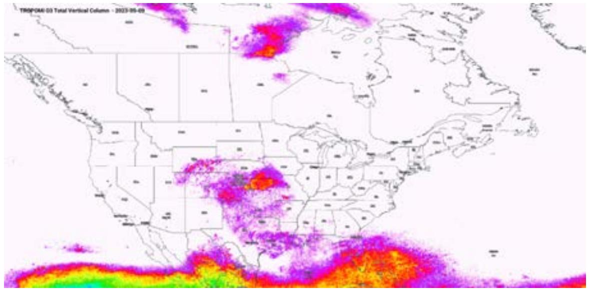
highlights – new and/or ongoing datasets (processing, image generation, ingestion)

Product	Type	Time Coverage	Location
PWWB Outputs (PM2.5, O3, NO, NO2, CO)	ML	2022-05-10 – ongoing	LA County
UCB Adjoint Sensitivities	Model	2010-12-01 – 2011-12-31	LA County / North America (NOx)
UCB Adjoint Emissions	Model	2010-12-01 – 2011-12-31	LA County / North America (NOx)
WUSTL PM2.5 Daily	Model	2017-07-02 – 2020-07-02	North America
WUSTL PM2.5 Monthly	Model	2017-01 – 2020-12	North America
GMU PM2.5 Hourly	Model	2017-10-01T14:00:00+0000 2018-12-18T23:00:00+0000	LA County
AERDB VIIRS/SNPP (Aerosols)	Satellite	2017-01-01 – 2018-12-31	Global
MERRA-2 Hourly (TLML, PBLH, CDH)	Satellite	2018-01-01T00:30:00+0000 2021-12-31T23:30:00+0000	Global
TROPOMI Daily (NO2, CO, O3, CH4, SO2)	Satellite	2020-07-13 – ongoing	California
TROPOMI Daily (NO2, CO, O3, CH4, SO2)	Satellite	2023-01-01 - ongoing	Global
TROPOMI Weekly (NO2)	Satellite	2021-12-31 - 2022-10-21	Ukraine
OMI Weekly (NO2, O3, SO2,	Satellite	2017-12-31 - 2022-10-28	Ukraine
OMI Monthly (NO2, O3, SO2,	Satellite	2004-10-01 - 2022-09-01	Global
GCHP Outputs (NO, NO2, SO2, O3, CH4)	Model	2017-03-01 – 2018-12-31	California
MODIS Aqua/Terra (Aerosols)	Satellite	2018-01-01 – 2023-04-24	Global
GEOS-CF	Model	In Progress	Global

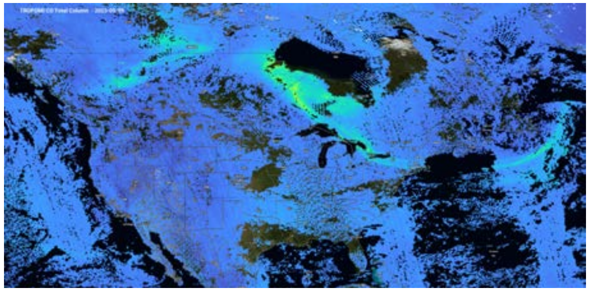
In Situ Data Products

Product	Type	Time Coverage	Location
GRIMM PM2.5	In Situ	2020-11-01T00:00:00+0000 2022-07-06T19:00:00+0000	JPL B301
Aethalometer Black Carbon	In Situ	2020-11-01T00:00:00+0000 2022-07-06T19:00:00+0000	JPL B301
PurpleAir PM2.5	In Situ	2022-07-01T00:00:00+0000 2022-07-07T21:00:00+0000	LA County
AQMD (PM2.5, SO2, CO, O3, NO2, N2O, CH4, CO2)	In Situ	2010-01-01T00:00:00+0000 2021-07-26T10:00:00+0000	LA County
Quant AQ (CO, CO2, NO, NO2, O3, PM1, PM2.5, PM10)	In Situ	2022-11-01T00:00:00+0000 Now	JPL B301
AirNow (PM2.5, O3, NO2)	In Situ	2023-05-01T00:00:00 2023-05-15T00:00:00	Alberta, Canada
AirNow (PM2.5, O3, NO2)	In Situ	2021-06-01T00:00:00 2023-05-15T00:00:00	California
Aeronet (AOD)	In Situ	In Progress	North America

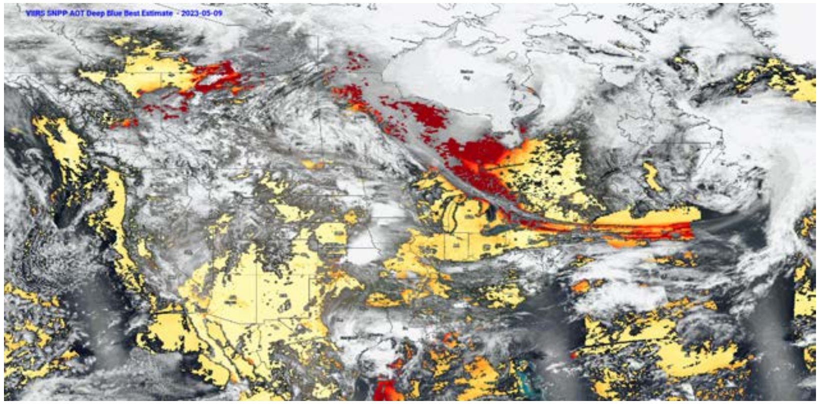
Visualize Canadian Wildfire Smoke



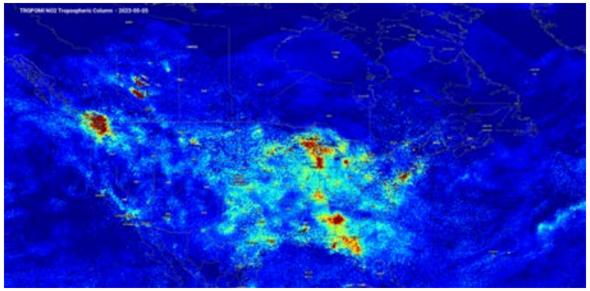
TROPOMI O3 Total Vertical Column
2023-05-09 – 2023-06-08



TROPOMI CO total column
2023-05-10 – 2023-06-09

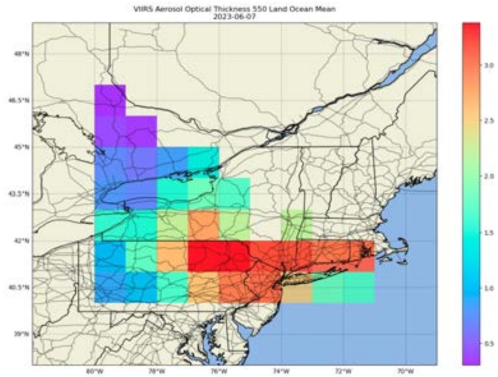


VIIRS SNPP Aerosol Optical Thickness Deep Blue Best Estimate
2023-05-10 – 2023-06-09

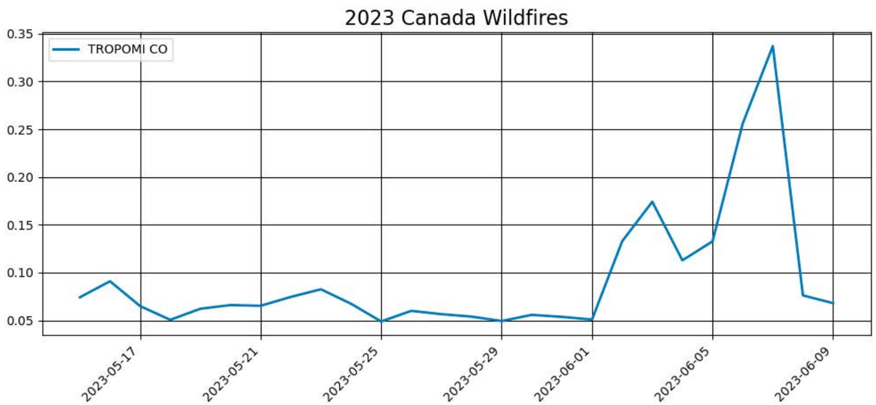


TROPOMI NO2 Tropospheric Column
2023-05-06 – 2023-06-05

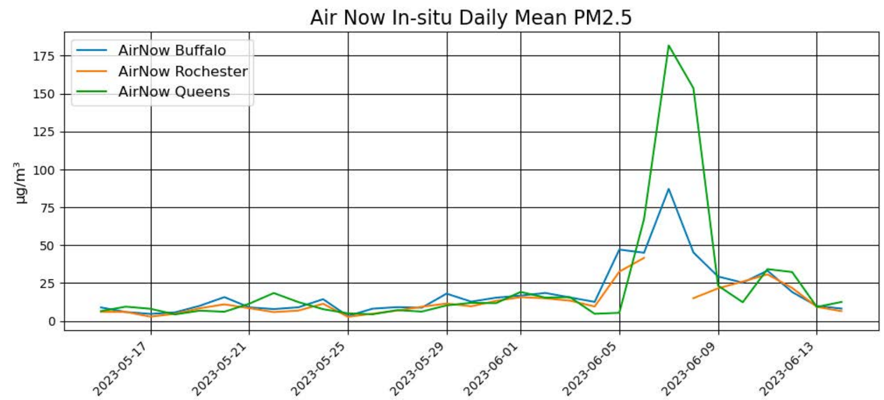
Analyze Canadian Wildfire Smoke



VIIRS Aerosol Optical Thickness
2023-06-07



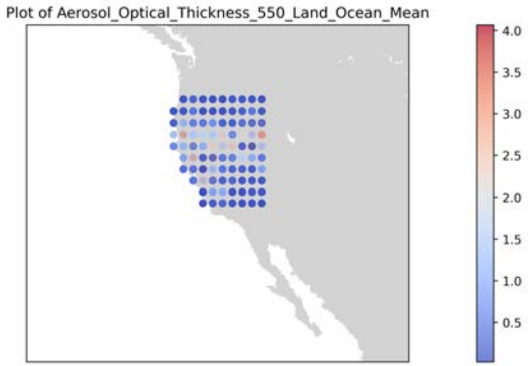
TROPOMI CO 2023-05-15 – 2023-06-09

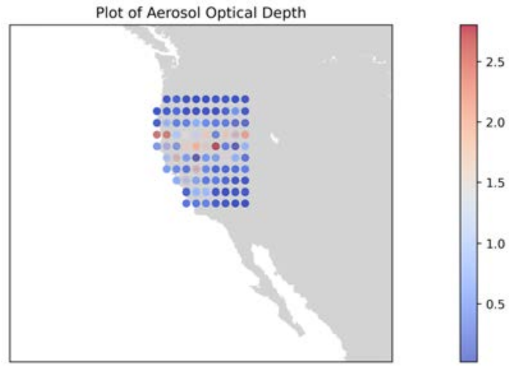


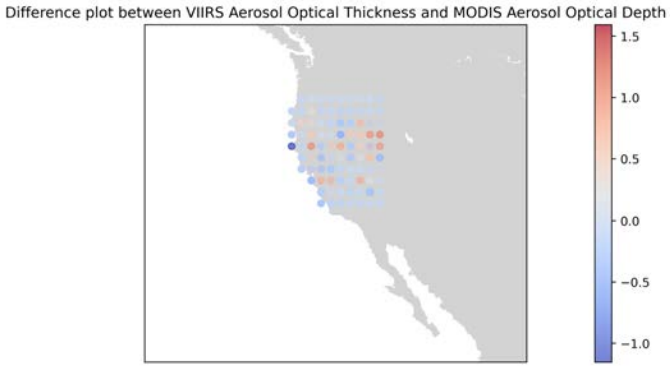
AirNow Buffalo, Rochester, and Queens $PM_{2.5}$ – 2023-05-15 – 2023-06-14

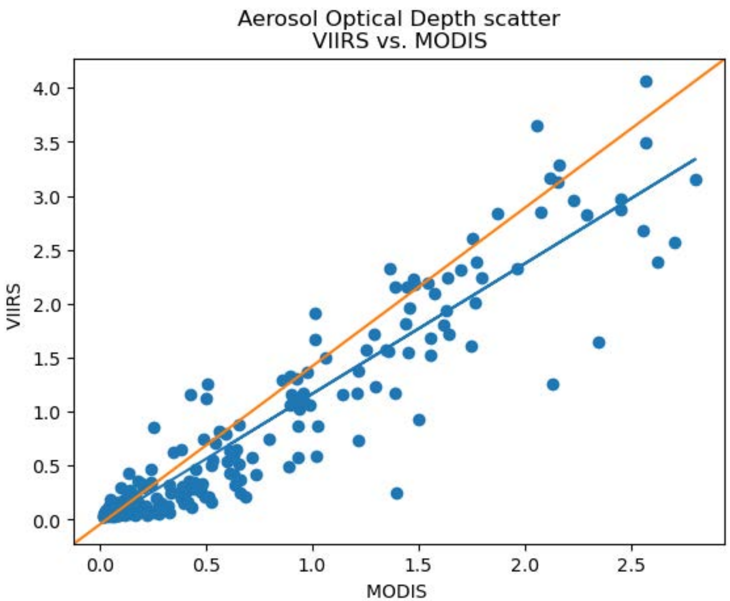
Satellite to Satellite Match-up – CA Wildfire Season

2021-08-07 to 2021-08-09



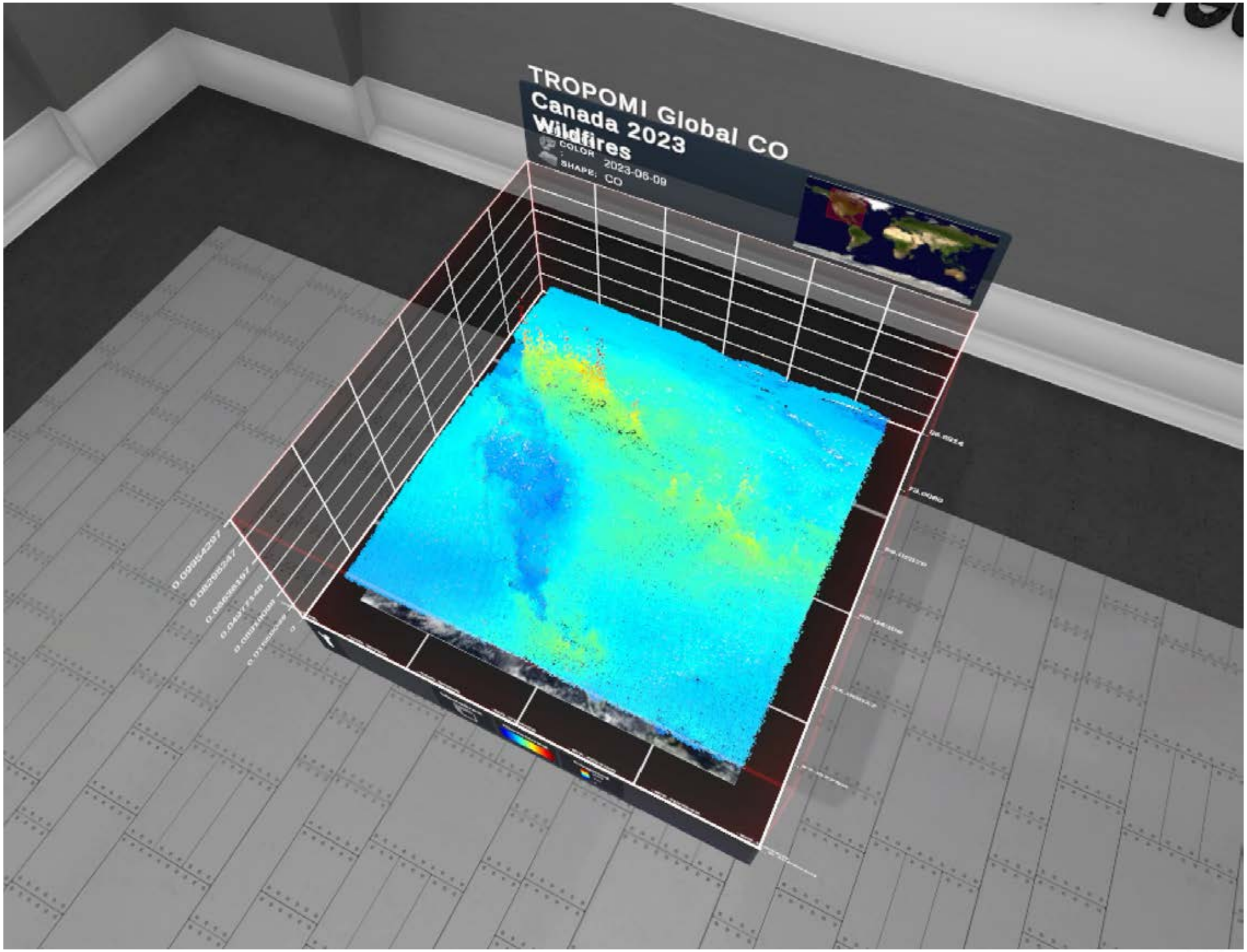




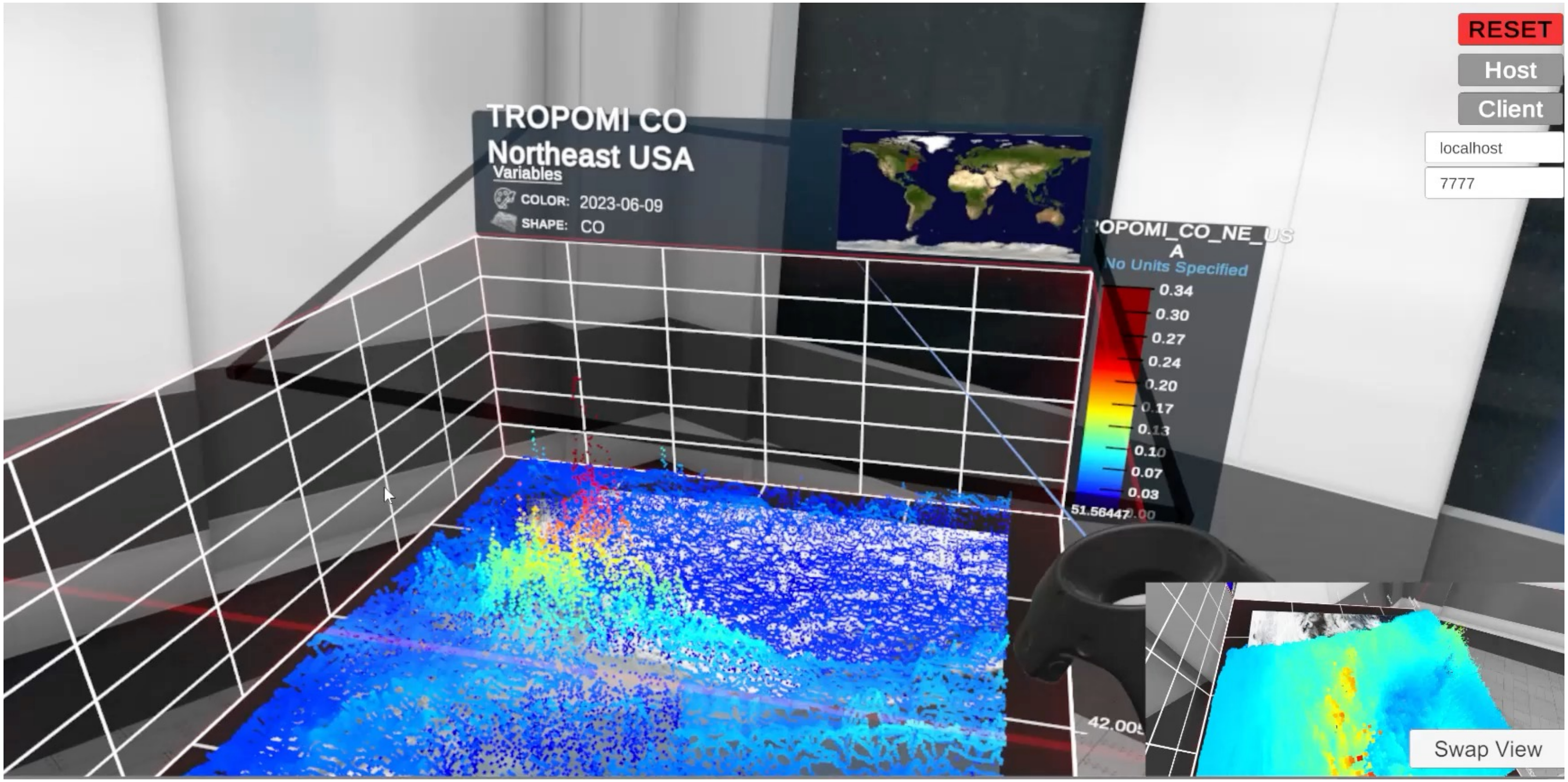


- Collaborated with NASA ACCESS Cloud-Based Data Matchup Service (CDMS) project to integrate matchup capability with Flood data holdings.
- Coordination between MODIS Aerosol Optical Depth and VIIRS Aerosol Optical Thickness

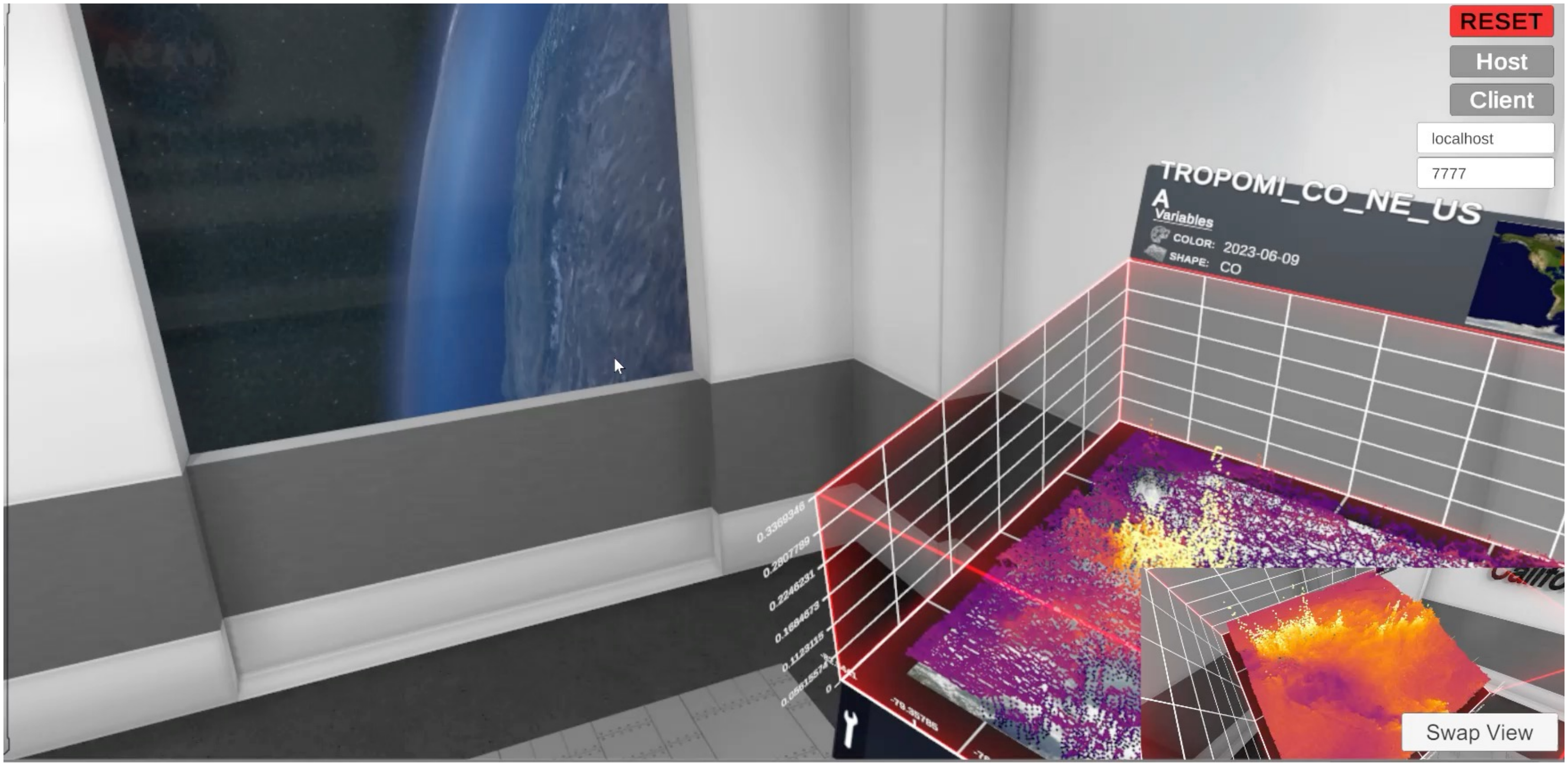
Immersive Science – Canadian Wildfire Smoke



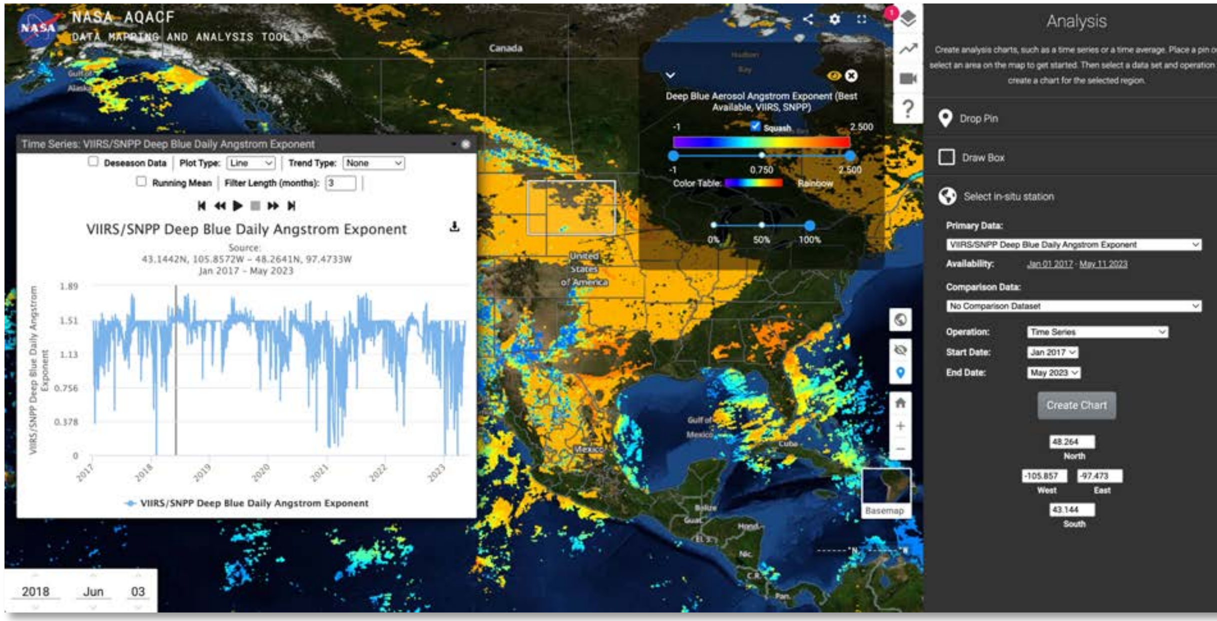
Immersive Science – Canadian Wildfire Smoke



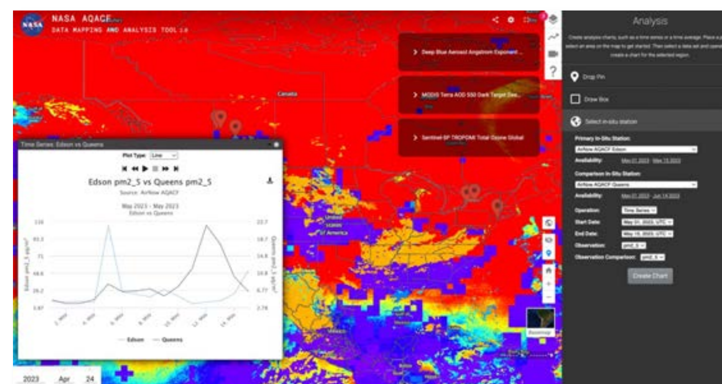
Immersive Science – Canadian Wildfire Smoke



Data Analysis Tool



- Users can now select in-situ stations and perform analysis on them as well as select their time bounds on the daily level
- Fixed colormaps and all colorbar features for layers (includes new legends, selectable palettes, and color range slider with squash option)
- Improved data layer selection behaviors
- Better data layer UI with three data categories: Satellite, Model, and AI
- Better default basemap with additional basemap options



New in-situ to in-situ comparison

ESTO Earth Science Technology Office

Air Quality Notebook

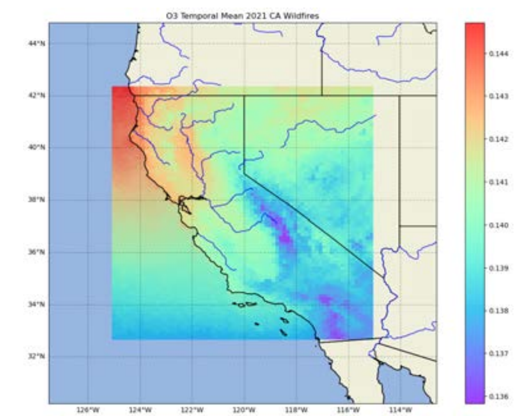
https://github.com/EarthDigitalTwin/FireAlarm-notebooks/blob/main/AirQuality_Demo.ipynb

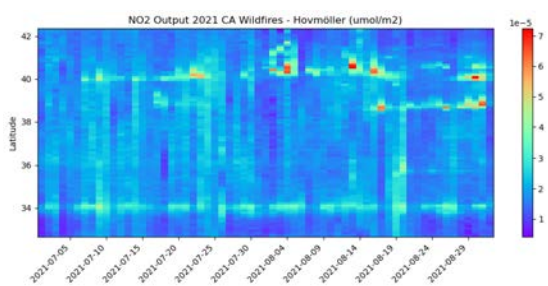
Demonstrates the latest IDEAS API for air quality analysis

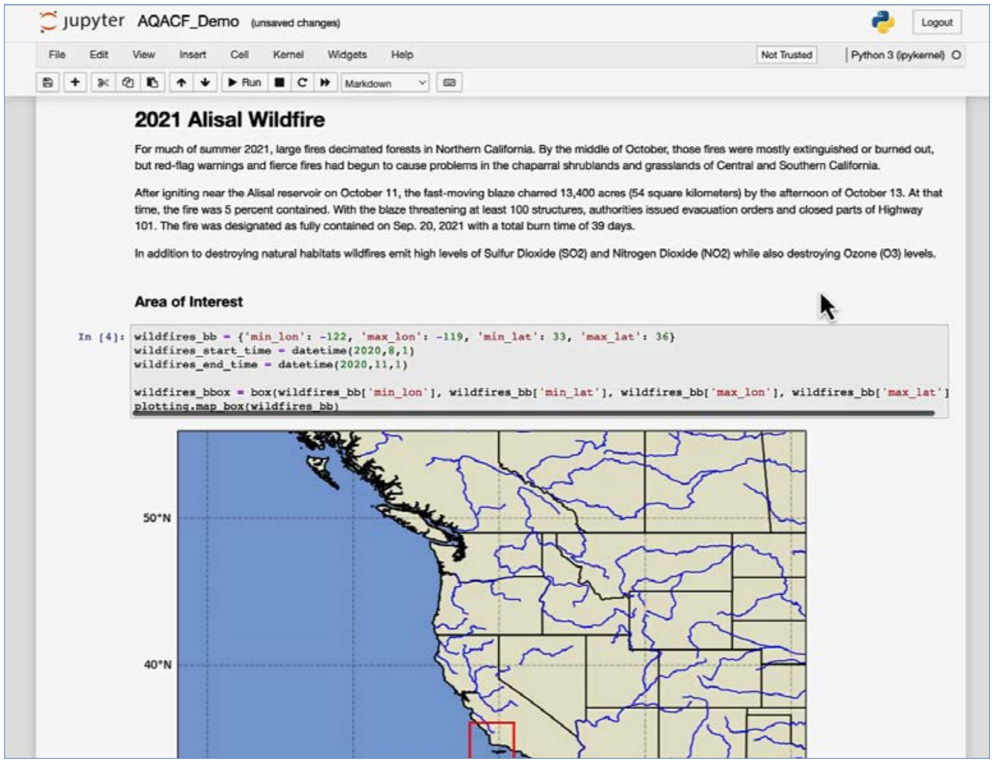
- STAC – Data search and metadata
- Data access – satellite, in-situ, and models
- Interactive, harmonized data analytic capabilities
- Visualizations – WMS and on-demand animation generation

Several use cases:

- 2021 Dixie Fire
- 2023 Alberta Canada Wildfires
- 2018 Carr Fire
- Los Angeles ports backlog Fall 2021
- Fireworks during 4th of July 2022 in Los Angeles county
- Predicting What We Breathe PM2.5 Predictions

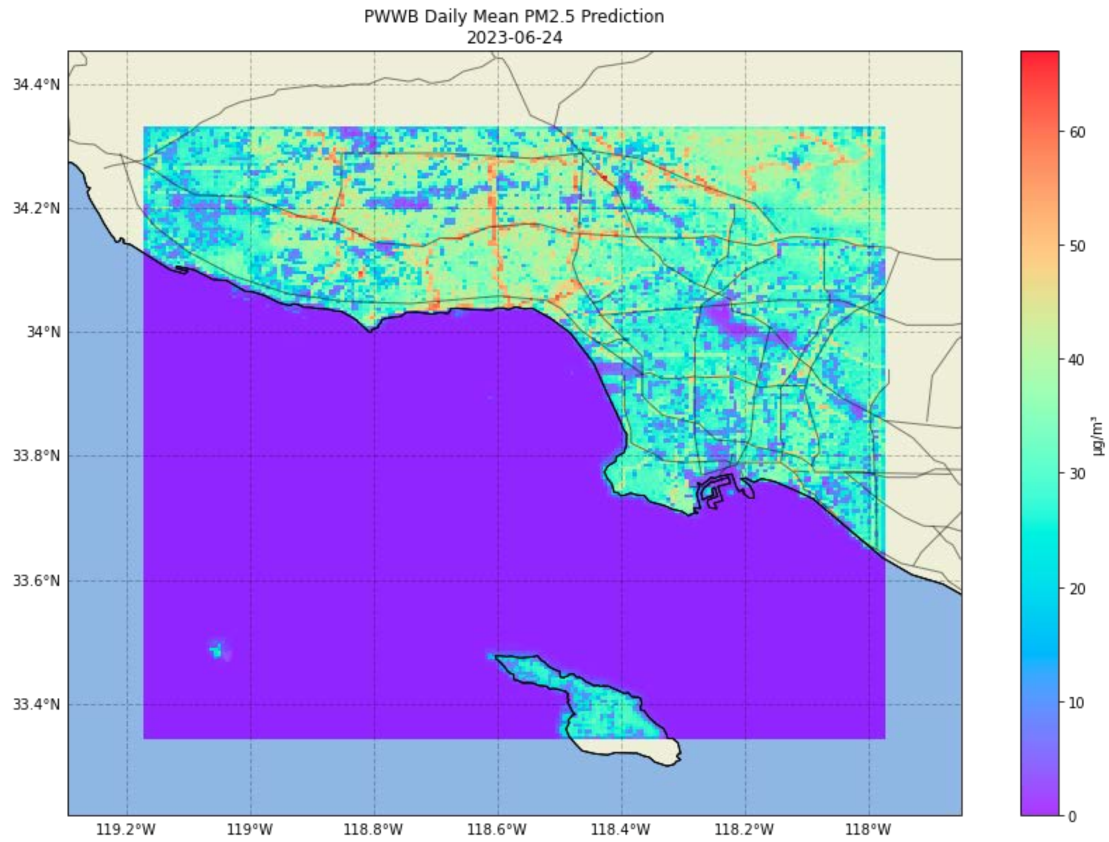






PWWB Integration

- Increased spatial resolution for multiple models ($PM_{2.5}$, NO_2, O_3)
- Incorporated Landsat 8 and 9 data as input for model training
- We encountered numerous integration issues on getting the PWWB models to produce consistent outputs
- The PWWB is looking into delivering their models in containers to streamline the integration



Machine Learning for In-Situ Data Acquisition and Retargeting

Purpose:

Perform large-scale (1000 km), multi-day prediction of air quality to prioritize in-situ data acquisition and MAIA target observations

Inputs:

VIIRS (Fire Detection), NOAA Hazard Mapping System (HMS) Smoke Levels, Wildfire Risk Index, Wind Information

Model:

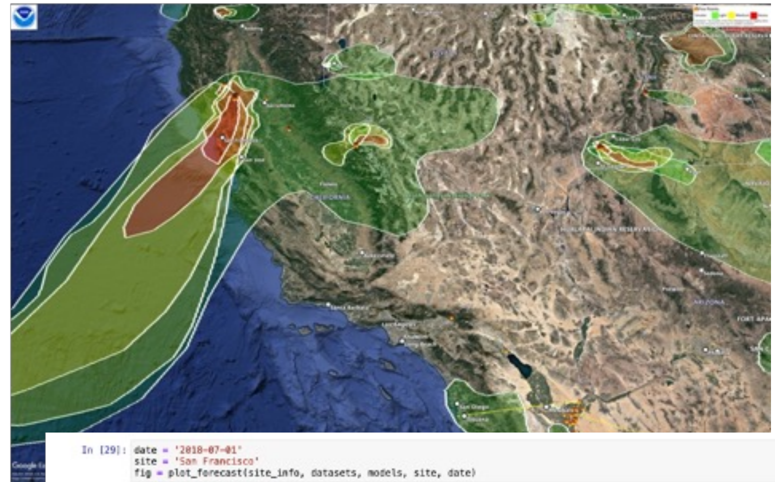
Predict smoke and pollutant levels at MAIA target sites and in-situ measurement sites for prioritization

Status:

- Prototype model using HMS data
- Identifying datasets for ingestion

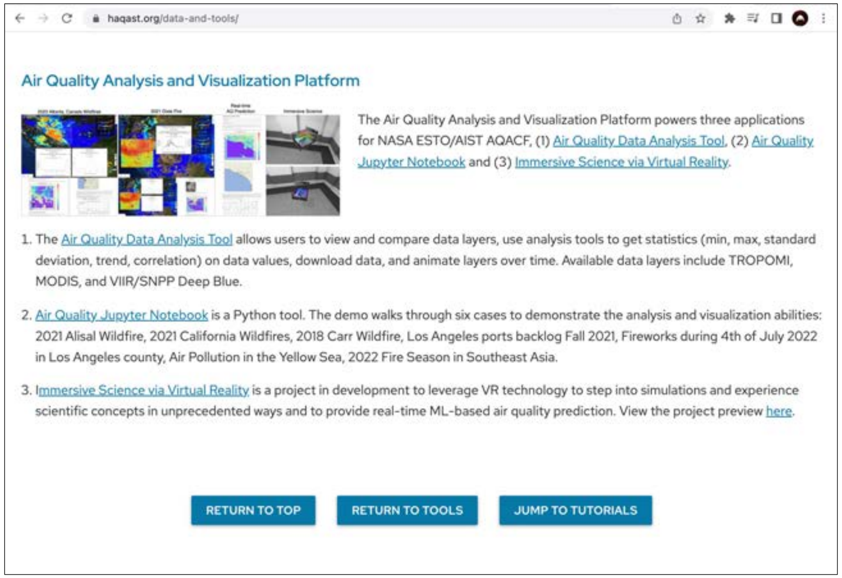
Next Steps:

- Ingest relevant datasets, improve model with additional inputs
- Integrate into data ingestion
- Demo MAIA observation planning scenarios



Recent Engagements and Highlights

- CEOS 55th Meeting of Working Group on Information Systems and Services (WGISS), Cordoba, Argentina
- CNES Toulouse Meeting
- EGU 2023, Vienna, Austria
- Group on Earth Observations (GEO) Dialogue Series on Open Software, Open Infrastructure, and Open Hardware
- EU's Standards in Digital Twins of the Ocean
- Upcoming
  - Oct. 19 and 20, 2023, HAQAST Public Meeting in Salt Lake City, Utah
  - IGARSS 2023 Hyperwall
  - IGARSS 2023 – "Application of Open-Source Digital Twins Framework for Wildfire and Air Quality"
  - ESIP Summer Meeting
  - Bulletin of the American Meteorological Society
  - Science Magazine Review Article



Fire Alarm / AQACF is now registered under NASA Health and Air Quality Applied Science Team (HAQAST)'s website

ESTO
Earth Science Technology Office

Q&A

EcoPro: Ecological Projection Analytic Collaborative Framework

Seungwon Lee (PI, JPL)
Peter Kalmus (Co-I, JPL)
Alex Goodman (Co-I, JPL)
António Ferraz (Co-I, JPL)
Flynn Platt (Co-I, JPL)
Emily Kang (Co-I, UC)
Jia Zhang (Co-I, SMU)
Kyle Pearson (Co-I, JPL)
Kyle Cavanaugh (Collaborator, UCLA)
Sudip Chakraborty (Collaborator, UMBC)

AIST-21-0032 Annual Technical Review
6/23/2023

1

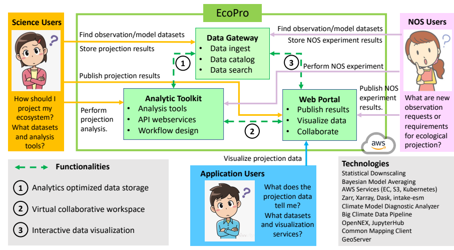
EcoPro: Ecological Projection Analytic Collaborative Framework

PI: Seungwon Lee, JPL

Objective

- Build **Eco**logical **Pro**jection Analytic Collaborative Framework (**EcoPro**).
- Perform scientific studies to demonstrate the scientific use of **EcoPro:** sequoia forests, kelp forests, and coral reefs.
- Generate application-usable datasets and visualizations to demonstrate the application use of **EcoPro**.
- Conduct experimental studies to demonstrate the NOS use of **EcoPro**.

EcoPro Architecture and Interface



Approach

- Develop (1) an analytic toolkit to perform the multidisciplinary analyses, (2) a data gateway to organize and access datasets, and (3) a web portal to visualize the study results and to provide a virtual collaborative work space.
- Leverage (1) CMDA and BCDP for the analytic toolkit, (2) AWS S3, zarr, and intake-esm for the data gateway, and (3) CMC, GeoServer, and OpenNEX for the web portal.
- Demonstrate the use of EcoPro in science, application, and NOS.

Co-Is/Partners: Kalmus, Goodman, Ferraz, Platt, Pearson (JPL), Zhang (SMU), Kang (UC), Cavanaugh (UCLA), Chakraborty (UMBC)

Key Milestones

• EcoPro requirements and architecture design	09/22
• EcoPro prototype (TRL 4)	02/23
• EcoPro science use case study report	07/23
• EcoPro application use case study report	12/23
• EcoPro NOS use case study report	05/24
• EcoPro end-to-end system (TRL 5)	07/24

TRL_{in} = 3 $TRL_{current}$ = 4

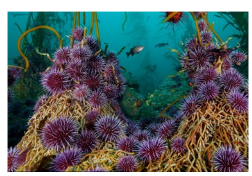
06/23 AIST-21-0032 2

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

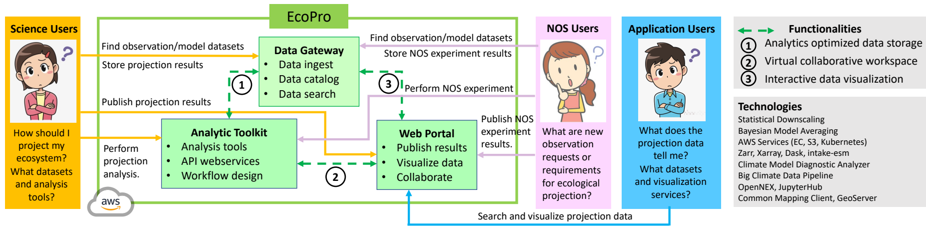
Background

- In this time of global heating and rapid climate change, Earth's ecosystems are under great stress for their survival and Earth's biodiversity is being rapidly reduced.
- Despite the importance of biodiversity for humanity and the imminent nature of the threat, efforts to project these losses over the coming decades remain crude.
- As a discipline, ecological projection is still in its early stage and will become increasingly important as stress drivers increase and losses mount.
- EcoPro will advance the ecological projection discipline by using cutting-edge data science methodologies to more optimally extract information from remote-sensing data, in-situ data, and ESM projections.
- EcoPro will help meet the R&A and Applications Science goals for Ecological Forecasting Program (Program Manager: Woody Turner).



Objectives

- Build **Eco**logical **Pro**jection Analytic Collaborative Framework (**EcoPro**): Analytic Server, Data Server, Collaboration Server, and Visualization Server.
- Perform scientific studies to demonstrate the scientific use of EcoPro: Giant sequoias forests, giant kelp forests, and coral reefs.
- Generate application-usable datasets and visualize them to demonstrate the application use of EcoPro.
- Conduct experimental studies to demonstrate the New Observing Strategies (NOS) use of EcoPro.



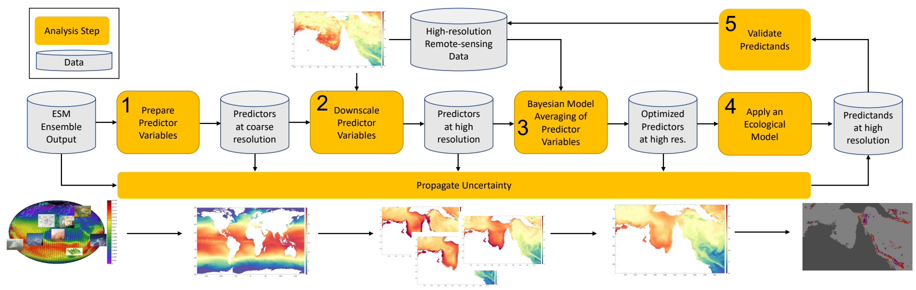
5

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

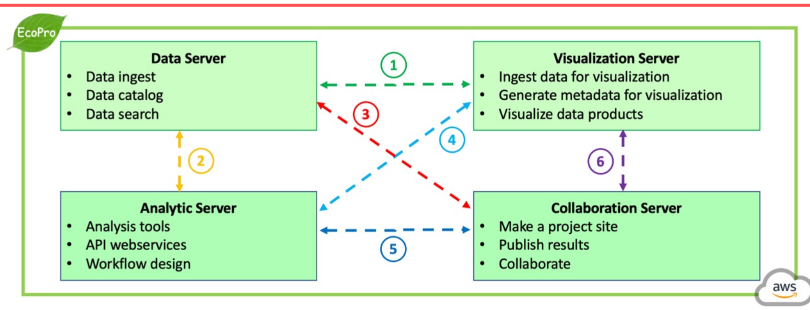
6

EcoPro Methodology Technical Advance



1. Collect, collate, and re-grid the ESM stressor/predictor variables;
2. Downscale the stressor variables using geostatistical inference and high-resolution remote sensing observations;
3. Skill-weight the ESM stressor variables against observations;
4. Apply the ecological model to perform the ecological projection and quantify the uncertainty of the projection;
5. Validate the ecological projections.

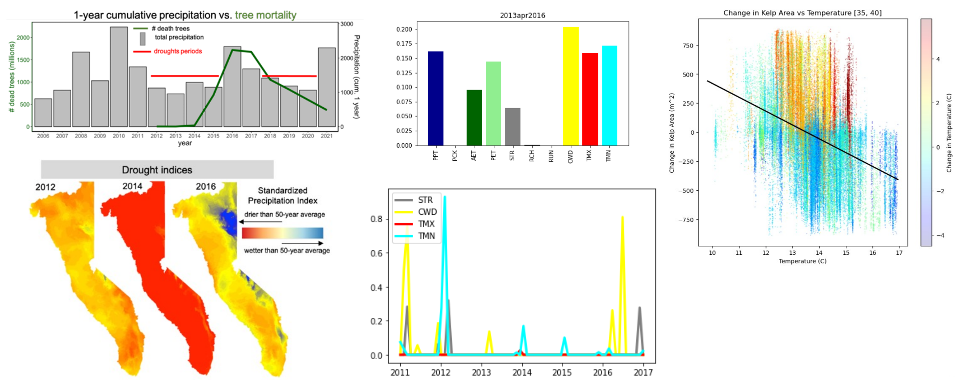
EcoPro System Technical Advance



- The Analytic Server enables big-data analytics with Xarray, Dask, Kubernates, Zarr, and BCDP (Big Climate Data Pipeline) tools.
- The Analytic Server supports the design of scientific workflows with reusable modules/components.
- The Data Server provides data ingesting, cataloguing, and searching services.
- The Visualization Server provides high-quality data visualization services for high-resolution ecological projection results.
- The Collaboration Server provides a virtual collaborative work space.

EcoPro Science Advance

- Harmonized the USFS tree mortality time series geodatabases to produce comprehensive time series (2012-2021).
- Developed drought indexes from monthly climate products.
- Identified important climate variables to model tree mortality.
- Found correlations between change in kelp area and sea surface temperature.

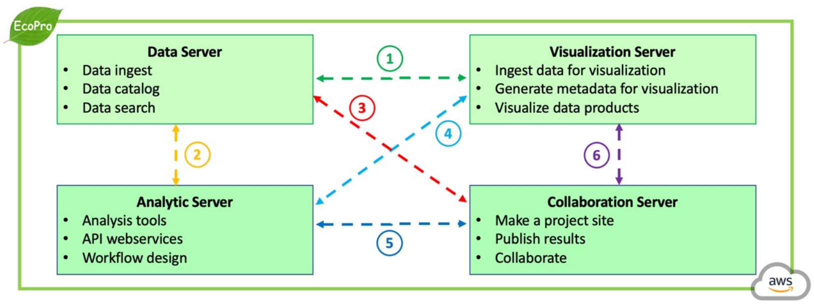


Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

Accomplishments

EcoPro System Architecture



Accomplishments

EcoPro System Requirements

Data Server:
- shall ingest data to the EcoPro database.
- shall store the ingested data in AWS S3 buckets.
- shall catalog data with meta data.
- shall support data search with meta data search conditions.
- shall convert large-continuous data into some cloud optimal data format Zarr.

Analytic Server:
- shall ingest analysis results back into data server for use by visualization server and web portal.
- shall maintain list of required third party libraries in a docker image.
- shall allow users to run data analysis workflows as Jupyter Notebooks.
- shall provide searchable interface for using ARD available from data server.

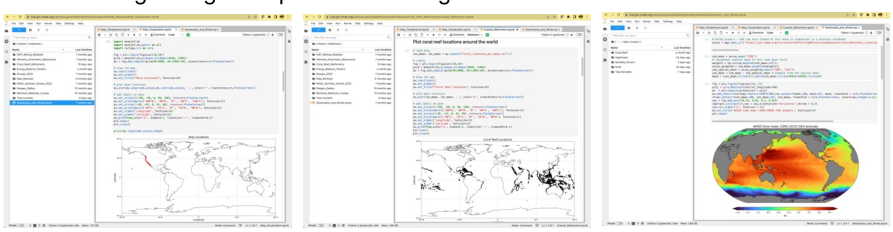
Visualization Server:
- shall ingest the EcoPro projection data products to the EcoPro visualization database (Data Ingester).
- shall generate the EcoPro visualization data products out of the EcoPro projection data products and store them in GeoServer.
- shall store the metadata and configuration data for each visualization data product in Config Server.
- shall visualize the visualization data products by retrieving configuration from Config Server and visualization data from GeoServer.

Collaboration Server:
- Shall provide a collaboration tool to communicate with a project team.
- Shall provide a project management tool to display the project overview, documents, and activities.
- Shall provide a publication tool to publish datasets, workflow tools, and other project results.
- Shall interface with the Data Server, Analytic Toolkit, and Visualization Server to link the project datasets, workflow, and result visualization.

Accomplishments

EcoPro Analytic Server

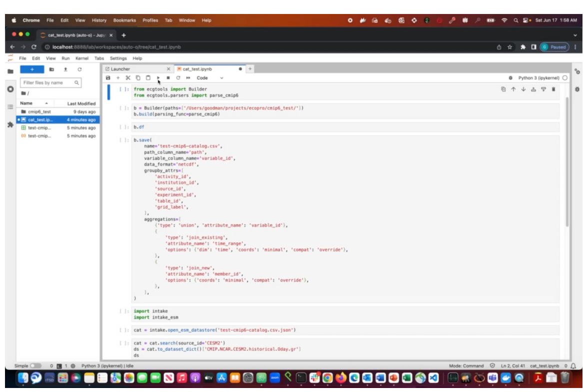
- Reused the JupyterHub server running on EKS from our previous AIST project for the EcoPro Analytic Server development
- Transferred cloud computing resources to us-west-2 AWS region.
- Made various software updates including setting up and configuring R studio/notebook kernel to allow development from team members in the R programming language.
- Reimplemented the following workflows on the EcoPro analytic server:
  - Seasonal subsetting for tree mortality
  - Regridding and spatial subsetting for CMIP6 SST



Accomplishments

EcoPro Data Server

- Developed prototype for ingesting/searching data processed from Analytic Server:
  - Catalog follows intake-esm json and csv file formats
  - Uses ecgtools to automatically keep all catalog files up to date
  - Full end-to-end workflow demonstrated for CMIP6 SST use-case



Accomplishments

EcoPro Visualization Server

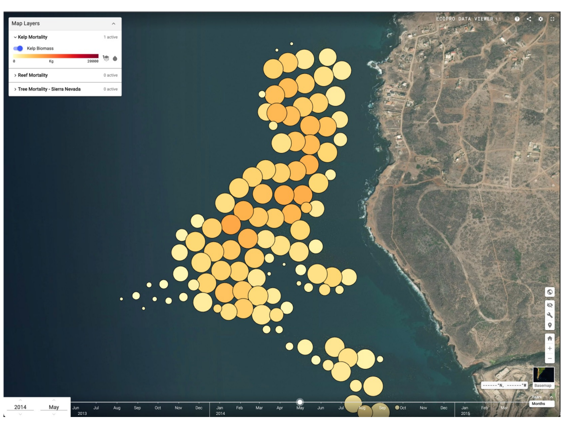
- Updated Common Mapping Client framework to build in modern development systems
- Ported the heritage CoralViz project to the updated framework
- Developed a backend application bundle for ingesting, serving, and viewing geospatial data.
    - Fully Dockerized and ready for cloud deployment
    - Python and Bash scripts for data processing
    - Support for GeoTIFF, NetCDF, and Shapefiles
- Extended the CMC base framework to include:
    - Dynamic pixel value extrapolation
    - Dynamic layer ingestion
    - Multi-layer configuration
    - Layer sub-grouping
    - Expanded vector data support
    - Additional time series controls
    - Dockerized container for build and deployment
- Developed a tool for generating useful raster color palettes.
- Processed, ingested, and visualized datasets for Tree Mortality, Reef Mortality, and Kelp Biomass

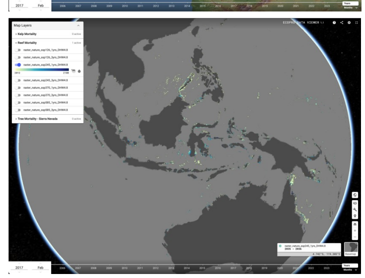
Earth Science Technology Office

15

Accomplishments

EcoPro Visualization Server





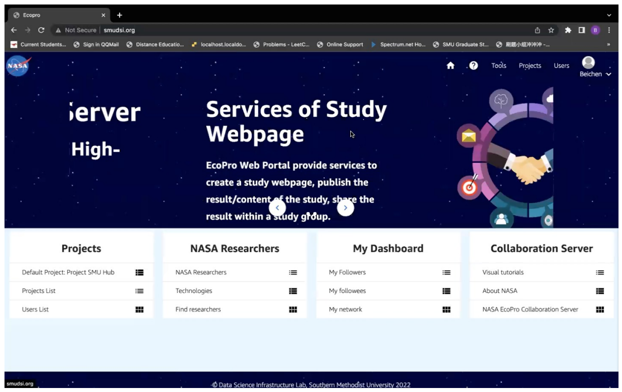
Earth Science Technology Office

16

Accomplishments

EcoPro Collaboration Server

- Developed, tested, and deployed the web-based collaboration portal to AWS.
- Implemented user access control and profile management functionalities.
- Implemented basic project publication and comment sharing functionalities.
- Implemented bug/suggestion/idea report and management functionalities.



ESTO Earth Science Technology Office

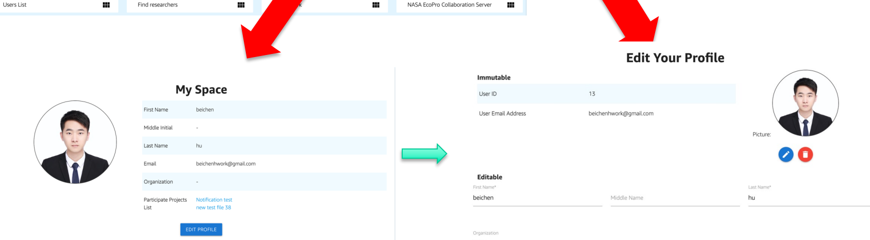
17

Accomplishments

Home Page

EcoPro Collaboration Server: Users

Users:
- Check user profile
  - Basic info
  - Participate projects
- Edit user profile
  - User personal info
  - Update user password



User Detail Page

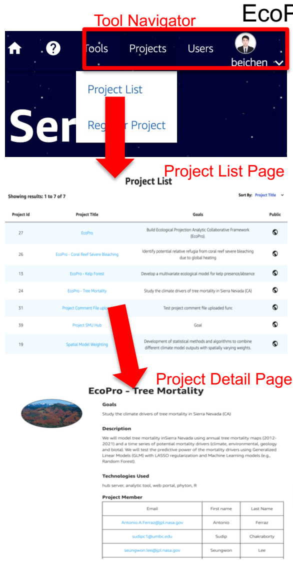
User Edit Page

ESTO Earth Science Technology Office

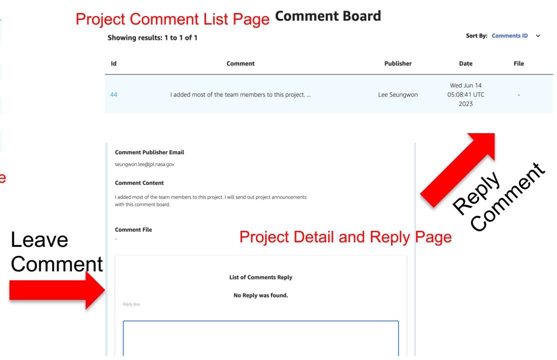
18

Accomplishments

EcoPro Collaboration Server: Projects



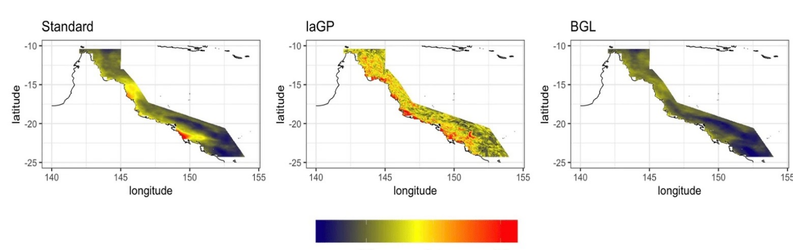
Projects:
- Check project detail
  - Goal, Description, Technology used, files…
- Comment project
  - Only group member visible
  - Reply comment and upload reply files



Accomplishments

Statistical Downscaling

- Completed the improved downscaling algorithms and validation with two different datasets, one with Gaussian distribution and the other with logistic transformation.
- Provided a brief user guide for the analytic server team to integrate the downscaling algorithms into the ecological projection workflow.



Comparison among different statistical downscaling methods. Our BGL method outperforms in terms of mean squared error against the MUR SST observational data in a central region of the Great Barrier Reef.

Accomplishments

Spatial Model Weighting

- For a given remote sensing data *Y* and a given set of climate model outputs *O*, find **the spatial model weighting factors ϕ_{ij}** as follows:

$$Y_i = \sum_{j=1}^{K} \phi_{ij} O_{ij}, \qquad i = 1, \ldots, n$$

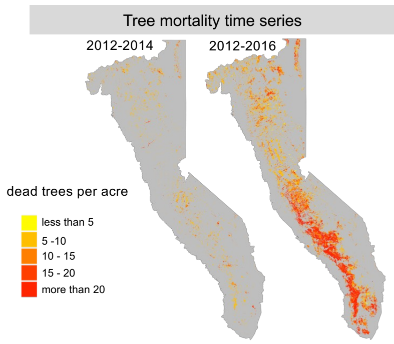
  - ***Y*:** *n*-dimensional vector of remote sensing data
  - ***O*:** *n* x *K* matrix whose columns are n-dimensional vectors of output from *K* climate models
  - *n* : the number of grid cells
  - $\sum_{j=1}^{K} \phi_{ij} = 1$, for $i = 1, \ldots, n$
  - $\phi_{1j}, \ldots, \phi_{nj}$ has a spatial dependence structure for $j = 1, \ldots, K$.

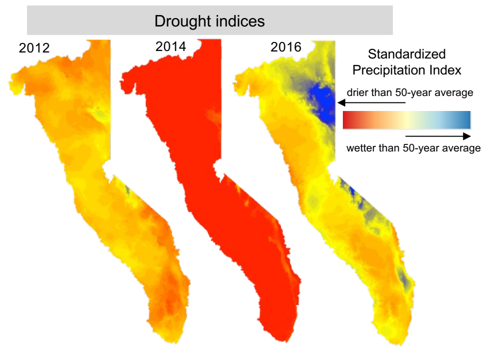
- Derivation of the MCMC algorithm is completed.
- The majority of the code written in R is completed.
- Will apply to the CMIP6 SST model outputs and MUR SST observation data.

Accomplishments

Tree Mortality Science Use Case

- Task plan to model tree mortality from environmental data (climate, soil moisture, topography), enabling predictions of tree mortality under future climate scenarios
- R scripts to harmonize the USFS tree mortality time series geodatabases to produce comprehensive time series (2012-2021).
- R and phyton scripts to process historical and future monthly climate products
- Developed drought indexes from monthly climate products

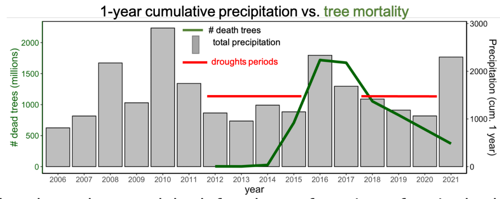




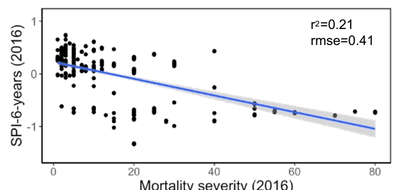
Accomplishments

Tree Mortality Science Use Case

- Analyses show that tree mortality manifested massively after 3 to 4 consecutive years of drought.



- Preliminary results show that models defined as a function of a single drought index encompassing multiple years (e.g., SPI 2012-2016) explain the peak of mortality in 2016.



- Next, we will develop feature-importance approaches to investigate further the primary climate drivers of tree mortality among multiple droughts indexes

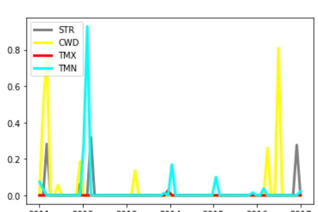
Accomplishments

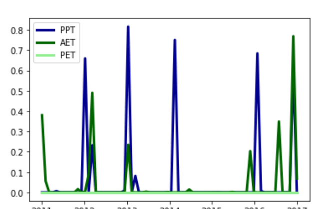
Tree Mortality Science Use Case

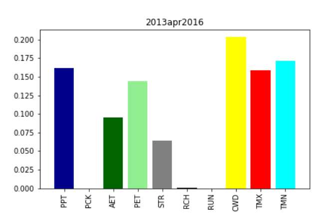
- Monthly data for ten different climate variables have been used to identify feature importance on tree mortality over the Sierra Nevada region.
- PPT, AET, PET, TMX, CWD, STR, and TMN are the dominant variables.
- PCK, RCH, and Run are the least important variables.
- PPT appears to be important every month.
- We estimated the correlation between the 7 major climate parameters and the tree mortality in year 2016 when the tree mortality was the highest.

Features

Precipitation (PPT)
Snow water Equivalent (PCK)
Actual ET (AET)
Potential ET (PET)
Soil water storage (STR)
Recharge (RCH)
Runoff (RUN)
Climate water deficit (CWD)
Maximum Temperature (TMX)
Minimum Temperature (TMN)



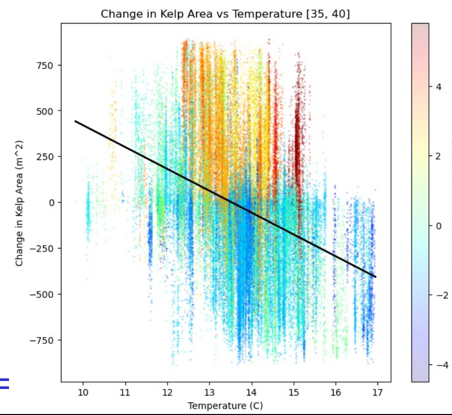


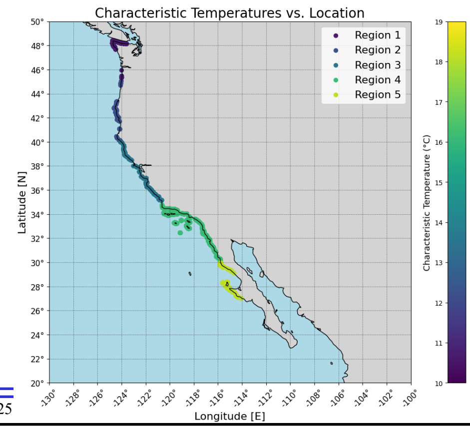


Accomplishments

Kelp Forest Science Use Case

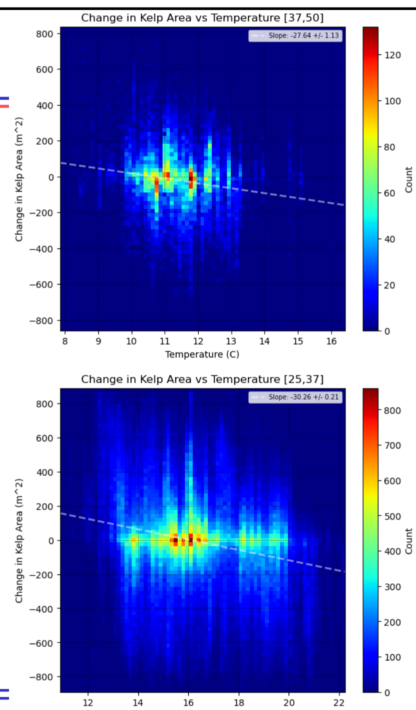
- Created multiple Jupyter notebooks to visualize, parse and analyze the bathymetric, sea surface temperature and kelp datasets
- Found correlations between the kelp presence/absence and sea surface temperature
- The characteristic temperature is defined when the change in kelp vs temperature is 0
- The next step is to explore forecasting models in the time domain





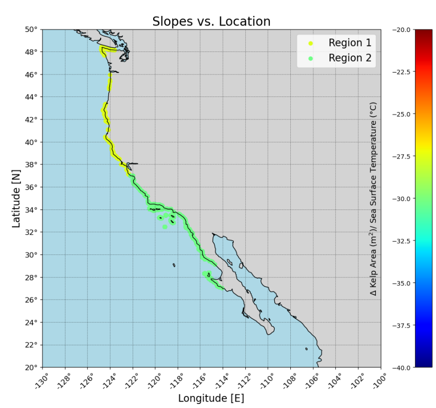
25

Accomplishments



Kelp Forest Science Use Case

- Estimated transition in kelp species around Half moon bay (~37 degrees)
- Northern response: -27.6+-1.1 m^2 /C (bull kelp)
- Southern response: -30.3+-0.2 m^2 /C (giant kelp)

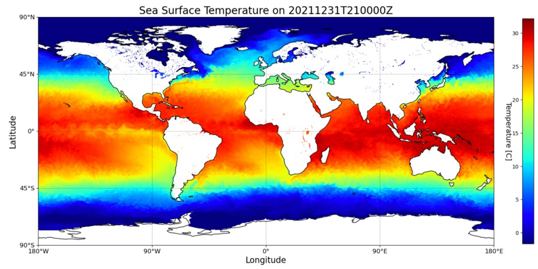


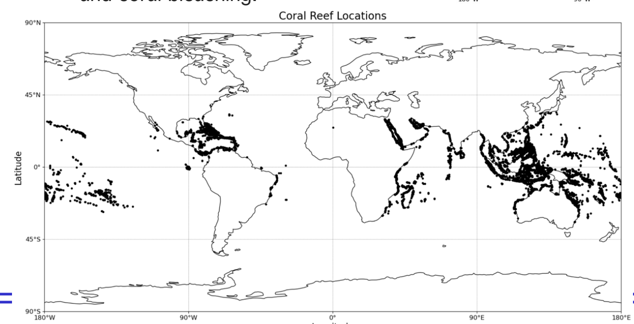
26

Accomplishments

Coral Reef Science Use Case

- Created multiple Jupyter notebooks to visualize, parse and analyze the bathymetric, SST and coral datasets
- Integrated data products from Allen Coral Atlas into EcoPro to track the growth and demise of coral.
- In the process of searching for correlations between SST, bathymetry and coral bleaching.





Plans Forward

- The Development Phase 1 of the Project is in progress (Mar 2023– Jul 2023).
  - Testing EcoPro Prototype with scientific use cases.
  - Developing the version 1 of EcoPro subsystems.
  - Developing an ecological projection model for kelp forest, tree mortality, and coral reef.
  - Producing statistical downscaled SST data products for kelp forest.
  - Generating harmonized data time series for tree mortality.
  - Generating climate indices for tree mortality.
  - Developing the model weighting algorithm.
  - Writing a workflow in Jupyter Notebook for ecological projections.
  - Milestone: EcoPro science use case study report.
- The Development Phase 2 of the Project is next (Aug 2023 – Dec 2023).
  - Will produce model-weighted kelp forest ecological projection results.
  - Will produce model-weighted tree mortality ecological projection results.
  - Will produce model-weighted coral reef ecological projection results.
  - Will integrate the statistical downscaling and model weighting code into Analytic Server.
  - Will integrate the ecological projection workflow into Analytic Server.
  - Will ingest and visualize the ecological projection results in Visualization Server.
  - Will connect Analytic Server and Visualization Server with Collaboration Server.

28

Presentation Contents

Actual/Potential Infusions and Collaborations

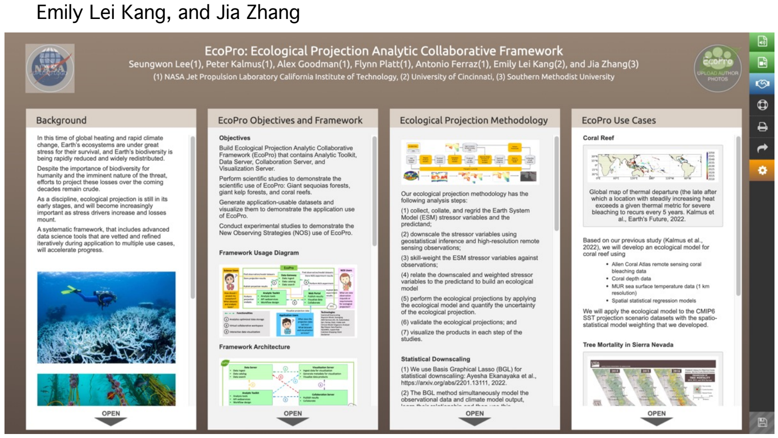
- Potential infusions
  - SBG science and application: will partner with SBG application lead Dr. Christine Lee and SBG science lead Dr. David Schimel in the second year to develop a new use case.
- Actual collaborations
  - Sudip Chakraborty at UMBC for tree mortality use case: Dr. Chakraborty provided the climate stressor datasets and working with us developing an ecological model.
  - Kyle Cavanaugh at UCLA for kelp forest use case: Prof. Cavanaugh provided the kelp mass datasets and technical support.
  - Collaboration with USFS resulted in new data products harmonizing time series of tree mortality surveys.
- Mentoring
  - JPL summer intern Nizhoni Sutter: mentoring her with the coral reef ecological model development task.
  - JPL summer intern Sierra Dahiyat: mentoring her with the kelp forest ecological model development task.

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Plans Forward
- Actual or Potential Infusions and Collaborations
- Publications - List of Acronyms

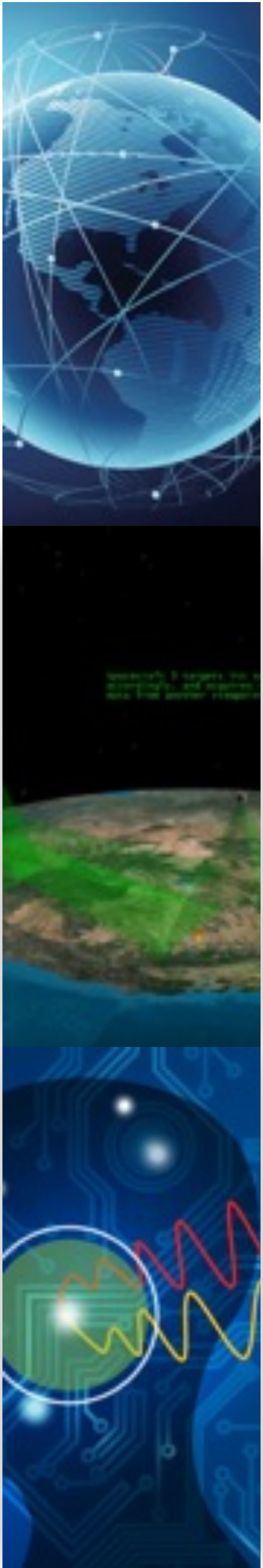
Publications

- AGU Fall Meeting Conference Poster (IN34A-03), December 12-16, 2022, Chicago, IL
  - Title: EcoPro: Ecological Projection Analytic Collaborative Framework,
  - Authors: Seungwon Lee, Peter Kalmus, Alex Goodman, Flynn Platt, Antonio Ferraz, Emily Lei Kang, and Jia Zhang



List of Acronyms

API: Application Programming Interface
AWS: Amazon Web Services
BCDP: Big Climate Data Pipeline
BMA: Bayesian Model Averaging
CMC: Common Mapping Client
CMDA: Climate Model Diagnostic Analyzer
ESM: Earth System Model
EcoPro: Ecological Projection Analytic Collaborative Framework
JPL: Jet Propulsion Laboratory
OCW: Open Climate Workbench
OIIP: Oceanographic In-Situ data Interoperability Project
OpenNEX: Open NASA Earth Exchange
PO.DAAC: Physical Oceanography Distributed Active Archive Center
R&A: Research and Analysis
SBG: Surface Biology and Geology
SMU: Southern Methodist University
SOTO: State of the Ocean (PO.DAAC's visualization tool)
TRL: Technology Readiness Level
UC: University of Cincinnati
UCLA: University of California, Los Angeles
UMBC: University of Maryland, Baltimore County
USFS: United States Forest Service



ESTO

Earth Science Technology Office