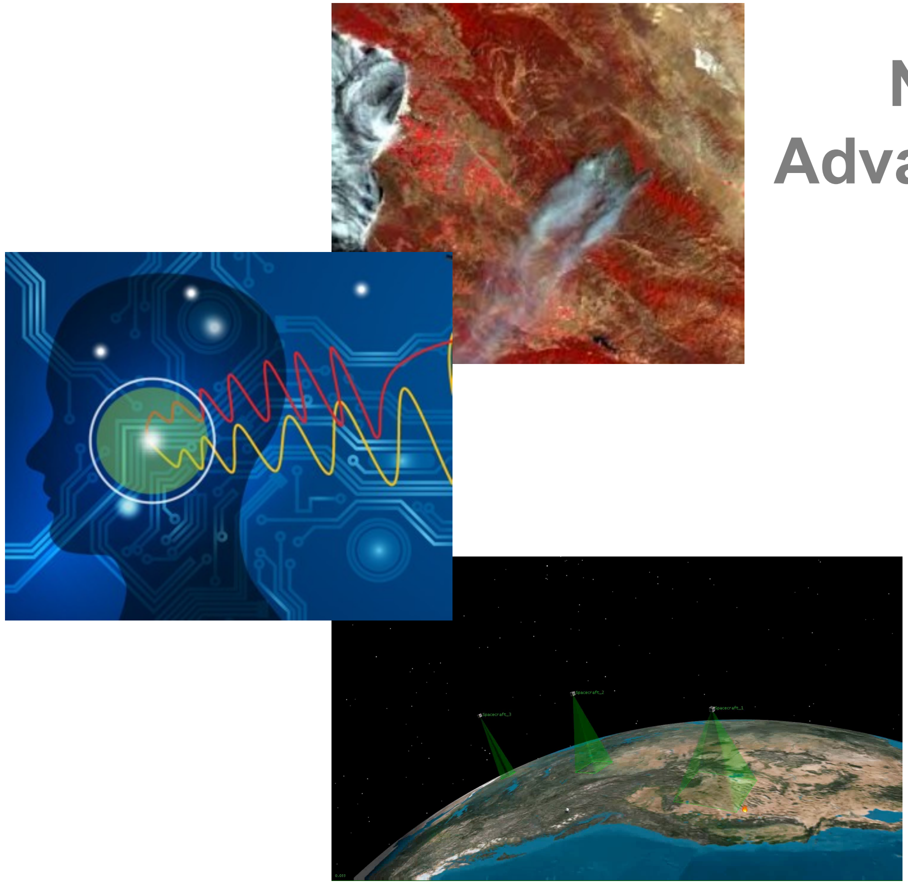


NASA Earth Science Technology Office (ESTO)
Advanced Information Systems Technology (AIST)

Analytics Collaborative Frameworks (ACF)

Annual Technical Reviews

Jacqueline Le Moigne

January 22 & February 5, 2021

Advanced Information Systems Technology (AIST) Program Management Team

“Investment in information systems that NASA Earth Science will need in the 5 to 10-year timeframe”

Jacqueline Le Moigne, Program Manager
Mike Seablom, Senior Strategist
Marge Cole, Outreach and Validation
<u>Associates:</u> Ian Brosnan, Transitions/Infusions Laura Rogers, Biodiversity & Ocean Nikunj Oza, AI & Knowledge Systems Ben Smith, Autonomy
Jackie Ferguson, Resources Analyst
Bob Connerton, Advisor
Paul Padgett, Communications

NOS and ACF for Science Data Intelligence

Optimize measurement acquisition using many diverse observing capabilities, collaborating across multiple dimensions and creating a unified architecture

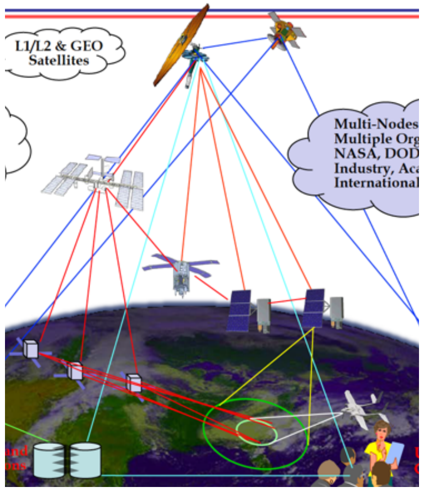
Assimilate Observations

Enhance and enable focused Science investigations by facilitating access, integration and understanding of disparate datasets using pioneering visualization and analytics tools as well as relevant computing environments

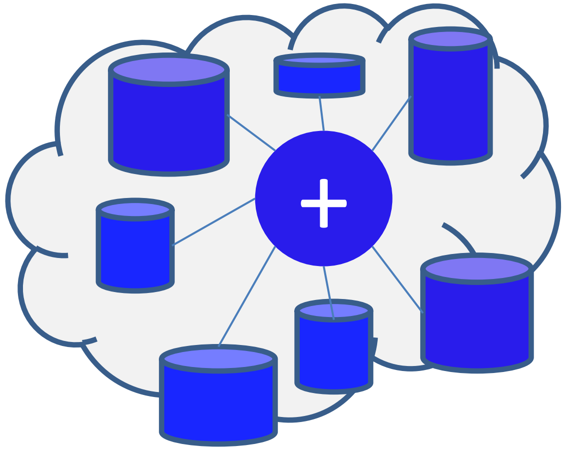
New Observing Strategies (NOS)

Acquire coordinated observations

Track dynamic and spatially distributed phenomena



Analytic Collaborative Frameworks (ACF)



Assimilate many various data into models and analytic workflows.

What additional observations are needed?

Example: NOS Testbed Demonstration planned for Spring 2021 targeting Mid-West Floods with LIS Models as well as Space and ground observations

Observation Requests

Example: OceanWorks, ACF for Ocean Science https://oceanworks.jpl.nasa.gov

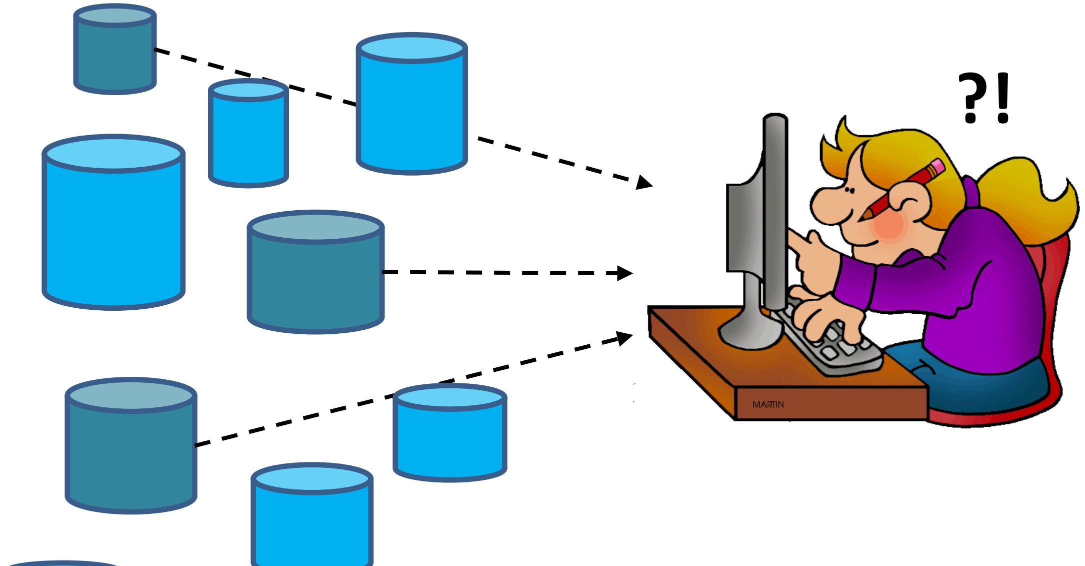
NOS+ACF acquires and integrates complementary and coincident data to build a more complete and in-depth picture of science phenomena

From Archives to Analytic Centers:
Focus on the Science User

Data Archives

Focus on data capture, storage, and management

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

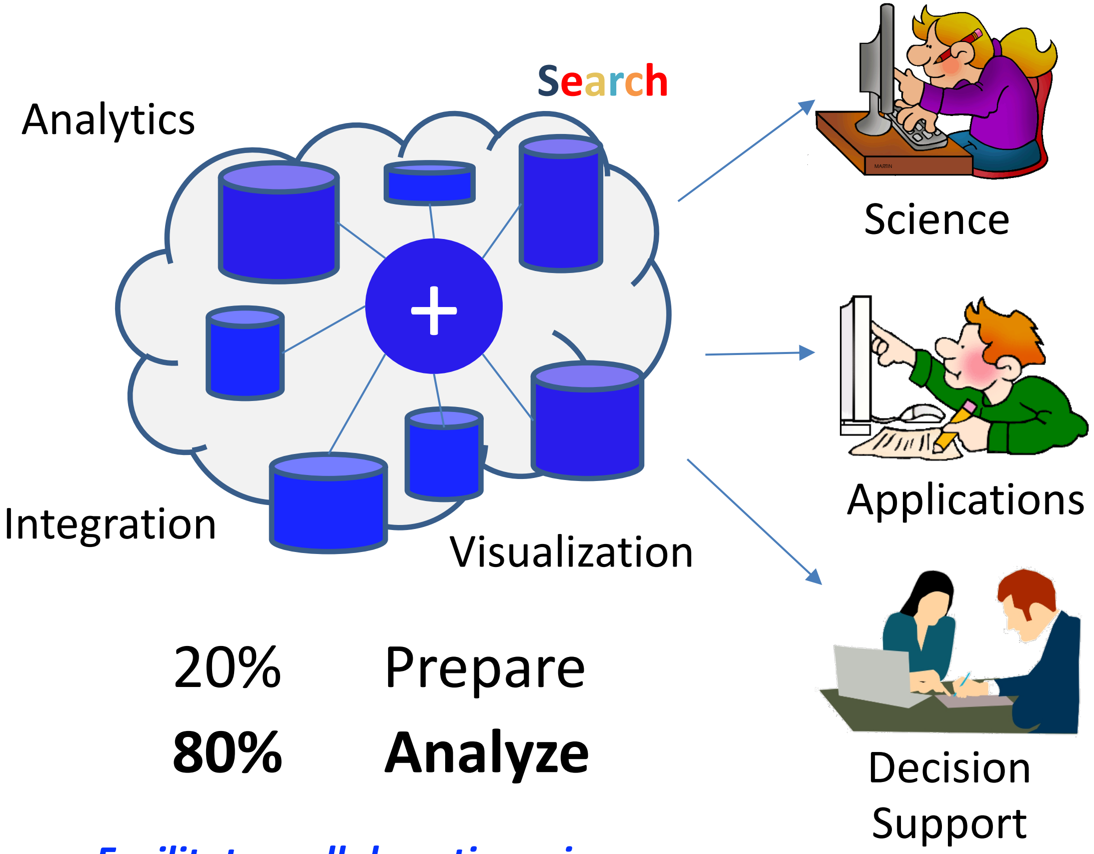


80%	**Prepare**
20%	Analyze

Analytic Centers

Focus on the science user

Integrated data analytics & tools tailored for a science discipline



20%	Prepare
80%	**Analyze**

Facilitates collaborative science across multiple missions and data sets

Analytic Collaborative Frameworks (ACF)

Focus is on the Science User

Allow flexibility/tailor configurations for Science investigators to choose among a large variety of datasets & tools

User
- Project Definition
- Plan for Investigation

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

Project Work Environment

Tools
- Discovery & Catalog
- Work Management
- Data Interfaces
- Analytic Tools
- Modeling
- Collaboration
- Visualization
- Sharing/ Publication
- Local/custom

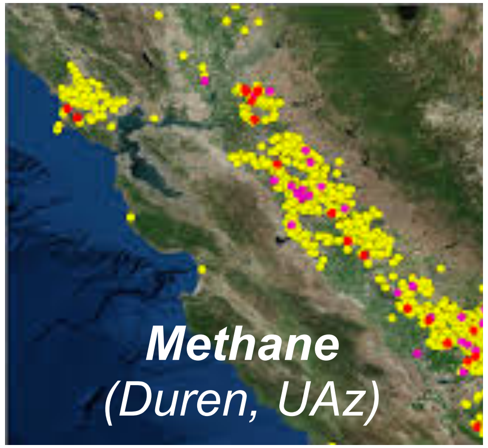
Reduce repetitive work in data access and pre-processing, e.g., develop reusable components

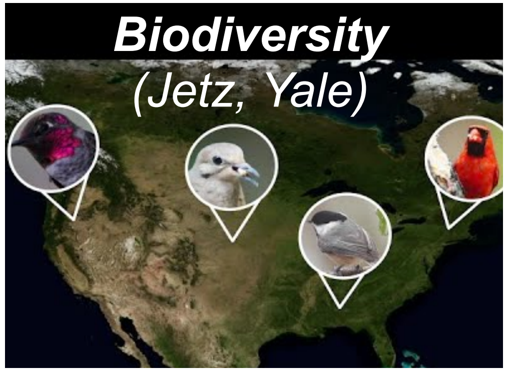
Computational Infrastructure
- Computing
  - Capacity
  - Capability
- Storage
- Communications

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

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

Analytic Collaborative Frameworks (ACF) support several Earth Science Disciplines





Technologies Currently Being Developed in ACF Projects

ADVANCED ANALYTICS:

- Data Accessibility (Duren, Jetz, Coen)
- Data Fusion (Donnellan, Duren, Jetz, Uz, Coen)
- Big Data Analytics (Hua, Ives, Swenson, Townsend)
- Data Mining (Donnellan)
- On-Demand Product Generation (Hua, Townsend)
- Data Operations Workflows (Zhang)
- Data Incorporation of Metadata, Provenance, Semantics, etc. (Huffer)

AI CAPABILITIES:

- Machine Learning (Beck, Holm, Huffer, Uz)
- Deep Learning (Beck, Holm, Huffer, Uz)
- Data Services Discovery (Zhang)
- Uncertainty Quantification Methods (Ives)

COMPUTATIONAL ENVIRONMENTS:

- Cloud Computing (Beck)

IMPROVED MODELING CAPABILITIES:

- Science Data Model Validation/Automation (Moisan)
- Science Code Development and Reuse (Henze, Moisan)
- Modeling Systems (Martin)
- Model Data Inter-Comparisons (Henze, Swenson)
- Custom Tools (Martin)
- Forecasting/Prediction (Jetz, Swenson, Townsend, Moisan)

ACF Review Schedule – 01/22/2021

January 22nd, 2021	Analytics Collaborative Frameworks (ACF-Group B) Technical Annual Reviews				
Tech	**Science**	**Name**	**Title**	**Start**	**Stop**
		Le Moigne	Introductions	11:00 AM	11:20 AM
Ceilometers, ML	PBL	**Halem**	A Deep Learning LIDAR-based Ceilometer Atmospheric Boundary Layer Height Over CONUS	11:20 PM	12:00 PM
Science Code Development, Model Data Inter-Comparisons	Atmospheric Composition, Atmos Gas	**Henze**	Surrogate modeling for atmospheric chemistry and data assimilation	12:00 PM	12:40 PM
Modeling Systems, Custom Tools	Atmospheric Composition, Atmos Gas	**Martin**	Development of GCHP to enable improved access to high-res atmospheric modeling	12:40 PM	1:20 PM
Autonomy, ML, Sensor Calibration & Validation	Atmospheric Composition, Total Ozone and Aerosols	**Holm**	Predicting What We Breathe: Using ML to Understand Urban Air Quality	1:20 PM	2:00 PM
		Break		***2:00 PM***	***2:20 PM***
Data Fusion, Data Mining	Earth Surface, Surface deformation	**Donnellan**	Quantifying Uncertainty and Kinematics of Earthquake Systems (QUAKES-A)	2:20 PM	3:00 PM
Big Data Analytics, On-Demand Products	Earth Surface, Surface deformation	**Hua**	Smart On-Demand of SAR ARDs in Multi-Cloud & HPC	3:00 PM	3:40 PM
Data Fusion & Accessibility	Carbon Cycle, Atmospheric Gas	**Duren**	Multi-scale Methane Analytic Framework	3:40 PM	4:20 PM
Data Operations Workflows, Data Services Discoverability	Climate variability, Global / regional climate systems	**Zhang**	Mining Chained Modules in Analytics Center Framework	4:20 PM	5:00 PM

ACF Review Schedule – 02/05/2021

February 5th, 2021 Analytics Collaborative Framewoks (ACF-Group A) Technical Annual Reviews					
Tech	**Science**	**Name**	**Title**	**Start**	**Stop**
		Le Moigne	Introductions	11:00 AM	11:20 AM
Data Fusion, Big Data Analytics	**Ocean Biology**	**Chirayath**	NeMO-Net – The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment	11:20 AM	12:00 PM
Autonomy, ML, Data Fusion	**Carbon cycle, ocean color**	**Schollaert Uz**	Shellfish aquaculture in the Chesapeake bay using AI for water quality	12:00 PM	12:40 PM
Science Data Modeling, Science Code Development	**Carbon cycle, ocean color**	**Moisan**	NASA Evolutionary Programming Analytic Center (NEPAC)	12:40 PM	1:20 PM
Autonomy, ML, Cloud Computing	**Rain Rate, Drop Size, Water & Energy**	**Beck**	Cloud-based Analytic Framework for Precipitation Research (CAPRi)	1:20 PM	2:00 PM
Big Data Analytics, Uncertainty Quantification	**Carbon cycle, Ecosystems**	**Ives**	Statistical tool to analyze large datasets for pattern changes and forecasting	2:00 PM	2:40 PM
		Break		***2:40 PM***	***2:50 PM***
Data Fusion, Data Accessibility	**Carbon cycle, Biodiversity**	**Jetz**	Biodiversity - Environment Analytic Center Modeling	2:50 PM	3:30 PM
Model Data Intercomparison, Big Data Analytics	**Climate variability, bio-diversity**	**Swenson**	Canopy condition to continental scale biodiversity forecasts	3:30 PM	4:10 PM
On-Demand Products, Big Data Analytics	**Carbon cycle, Biodiversity**	**Townsend**	GeoSPEC	4:10 PM	4:50 PM
Autonomy, ML, Metadata	**Carbon cycle, Ecosystems**	**Huffer**	AMP: An Automated Metadata Pipeline	4:50 PM	5:30 PM

AIST Group Project Review Objectives

- **Regular Annual Reporting Requirements**
  - Individual Programmatic Annual Reviews
  - Technical Annual Reviews Grouped by Topics
- **Establish relationship between awardees**
  - Introduce AIST PIs and their work to one another
  - Enable desired collaborations
  - Potentially share algorithms, codes or cross-cutting ideas
  - GoogleDocs:
    https://docs.google.com/document/d/1CvmgehHflwqDoTKtmrq7bdCm7NMY30bh1u2cHpIv5g8/edit?usp=sharing
- **Present AIST-18 Projects and PIs to broader community**
  - Present AIST-18 projects to NASA ESD Program Managers and partner organizations
  - Support technology infusions and knowledge transfer of AIST projects upon completion.
- **Review Needs in terms of:**
  - ESIP: Project analysis to improve infusion and transition opportunities
  - SMCE (NASA Science Managed Cloud Environment): AWS system access

ESIP
Earth Science Information Partners

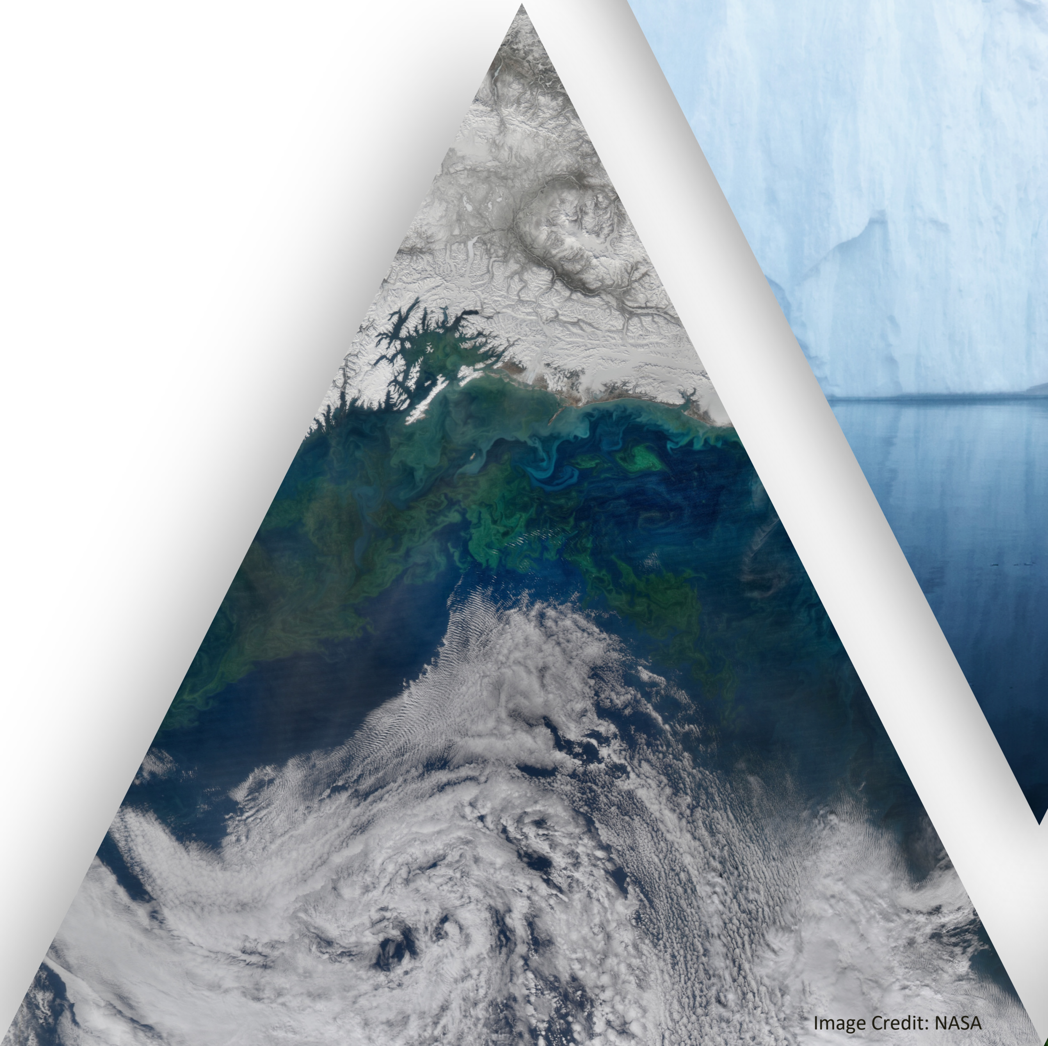


Image Credit: NASA

Image Credit: National Geographic

ESIP TECH EVALUATION

ANNIE BURGESS, PHD

AIST | Jan 22, 2021

Supported by:

ESIP PROVIDES AN EVALUATION FRAMEWORK THAT EXPOSES DEVELOPING TECHNOLOGY TO POTENTIAL END-USERS AND ADOPTERS, ULTIMATELY INCREASING ITS UTILITY AND USABILITY.

FUNDING

Evaluators are compensated for their time, increasing the likelihood of a thorough, comprehensive evaluation.

FACILITATION

ESIP facilitates evaluator calls, development of evaluation plan, communication with PIs.

FRAMEWORK

OBJECTIVES

ESIP works with PIs to set specific objectives taking into consideration TRL.

TECHNICAL EXCHANGE MEETING

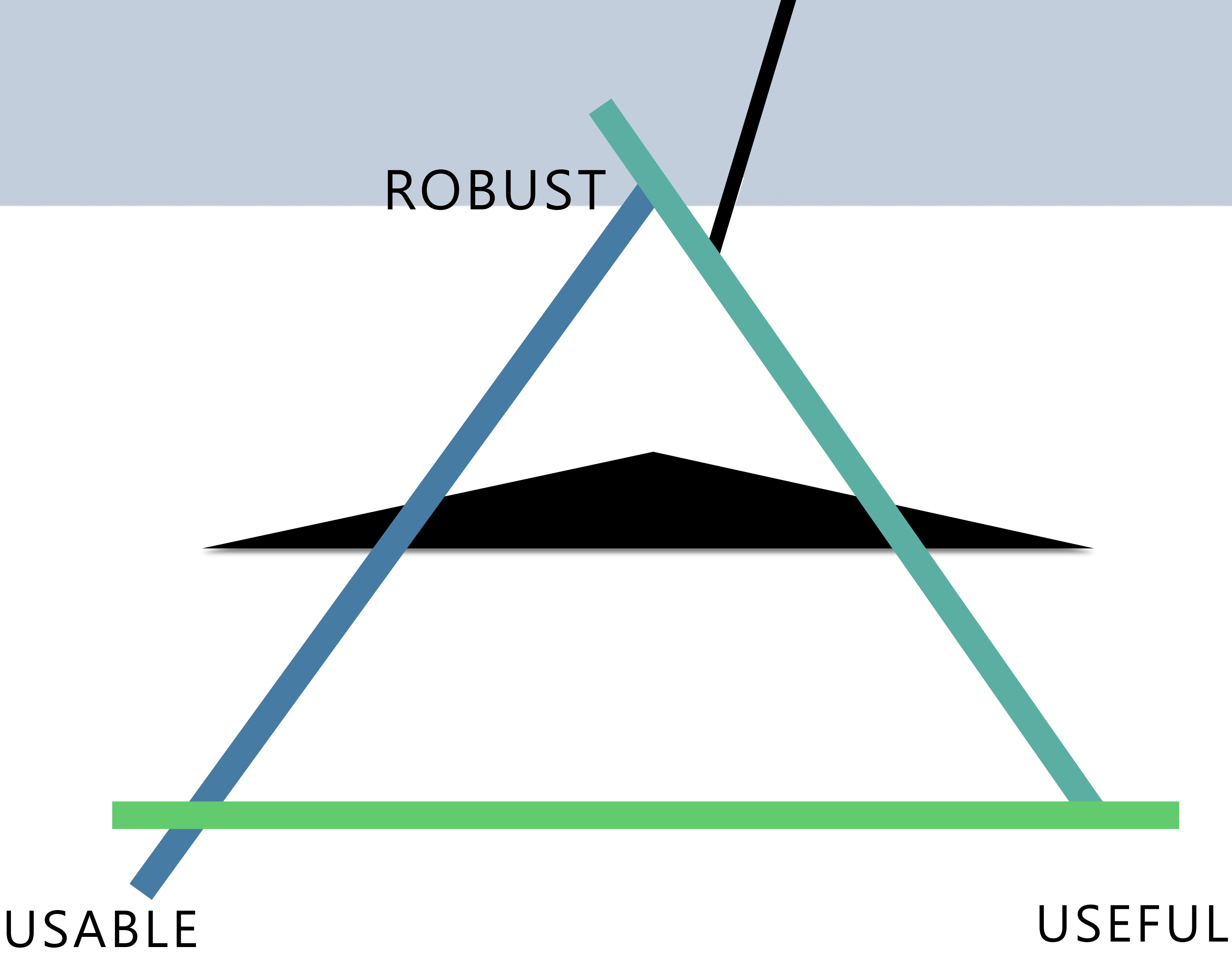
PI team meets evaluators. Big picture to backend... evaluators should have a solid understanding of the purpose and goals of tech.

EVALUATION PERIOD

ESIP coordinates evaluation process. Evaluators meet regularly, requesting information from PIs when necessary.

FINAL REPORT

ESIP works with evaluators to create final report to be shared with PIs & AIST. Reports can be public upon PI request.

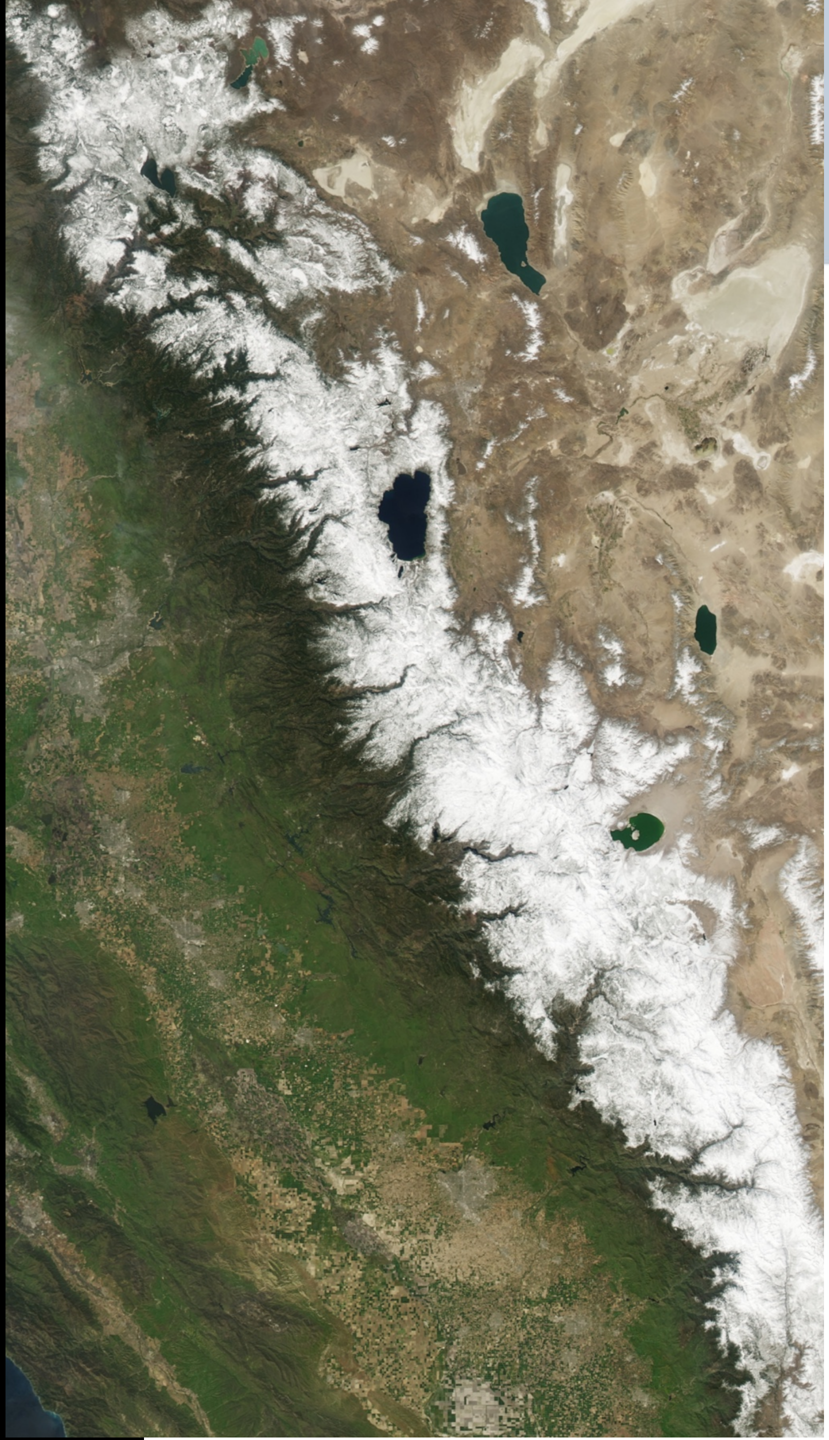

ROBUST
USABLE
USEFUL

ESIP
Earth Science Information Partners
esipfed.org | #ESIPfed

THANK YOU

ANNIE BURGESS, PHD

ANNIEBURGESS@ESIPFED.ORG



ESIP is supported by:

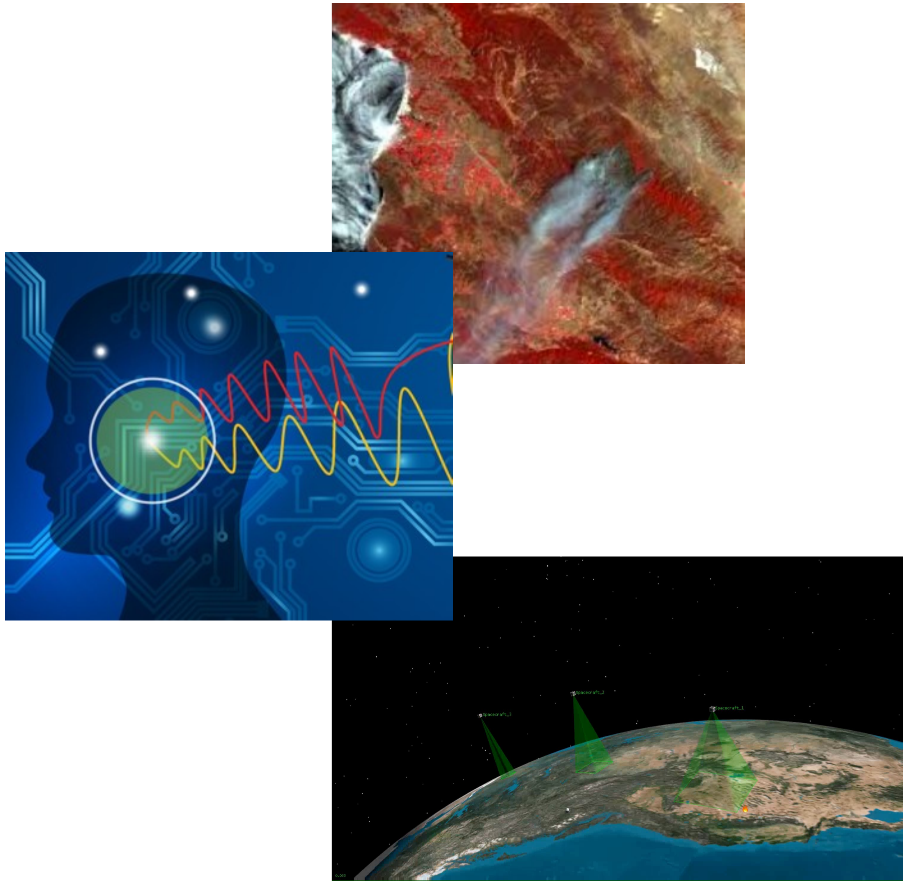
AIST SMCE Options
Marge Cole

- A critical component of the success of AIST projects is access to cost effective, flexible, and scalable compute and storage infrastructure.

- The Science Managed Cloud Environment (SMCE) is a managed Amazon Web Service (AWS) based infrastructure for NASA funded projects that can leverage cloud computing capabilities. This environment is designed to:
  - Provide cloud access to NASA PIs with non-NASA team members.
  - Perform research using new computing capabilities without extensive start-up time.
  - Use new tools and methods from AWS's product catalogue easily and affordably.
  - Scale computing for high-demand, high-bandwidth needs.

- More information at: https://www.nccs.nasa.gov/systems/SMCE

- ***NASA Managed (AWS) Cloud Environment Access***
  - Pay-as-you-go cloud account access with NASA security already built in
  - Enables ease of cloud-based project transition to NASA programs due to NASA level security already requirements already being met.



PI's Introductions

Around the Virtual Room


NASA
NEMO
NET
NEMO-NET - THE FLUID LENSING NEURAL NETWORK FOR GLOBAL CORAL REEF ASSESSMENT
AIST PROJECT REVIEW - FEB 2021
VED CHIRAYATH, ALAN LI, MICHAL SEGAL-ROZENHAINMER, JARRETT VAN DEN BERGH, JUAN TORRES-PEREZ, SAM PURKIS, SYLVIA EARLE
NASA AMES
LABORATORY FOR ADVANCED SENSING
NASA

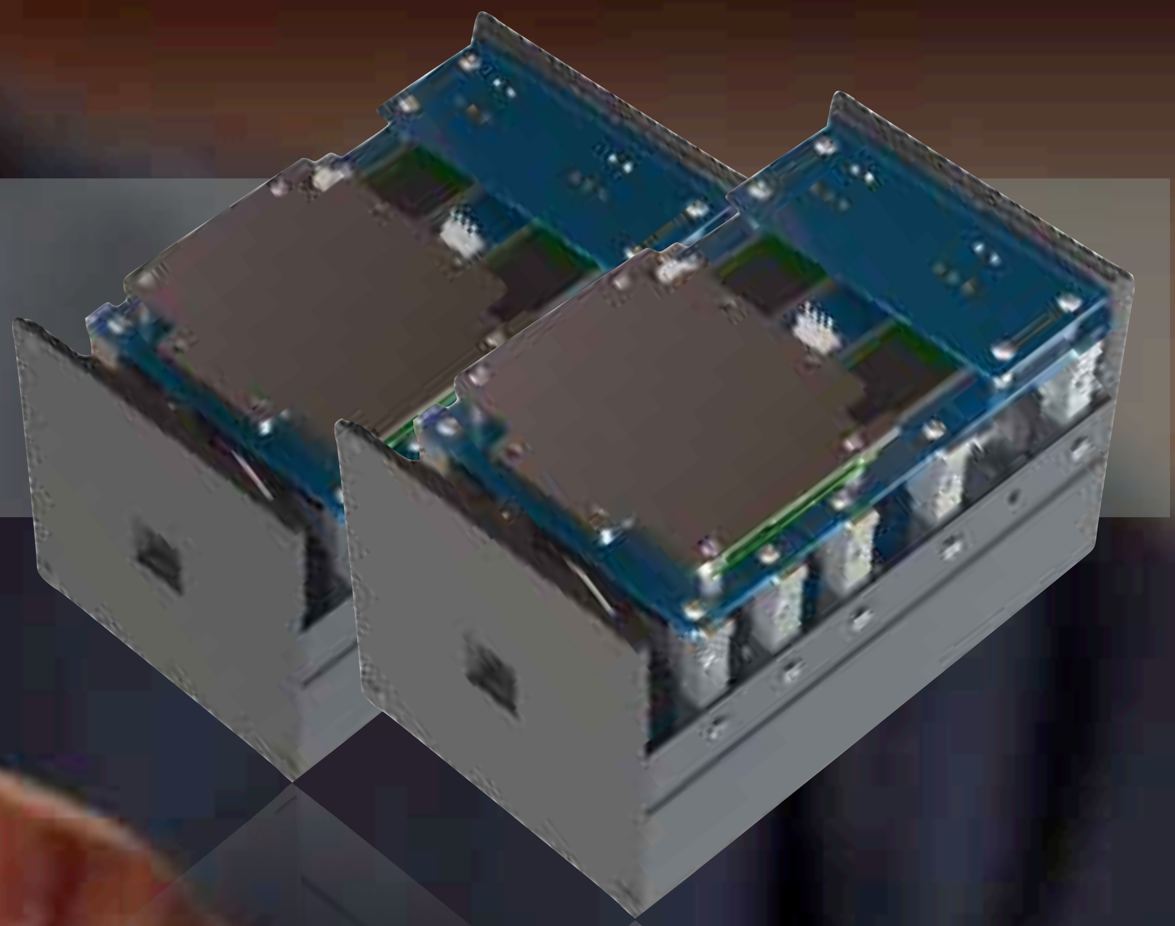

SCUBAPRO


National Aeronautics and Space Administration
Earth Science Technology Office
2020 Annual Report

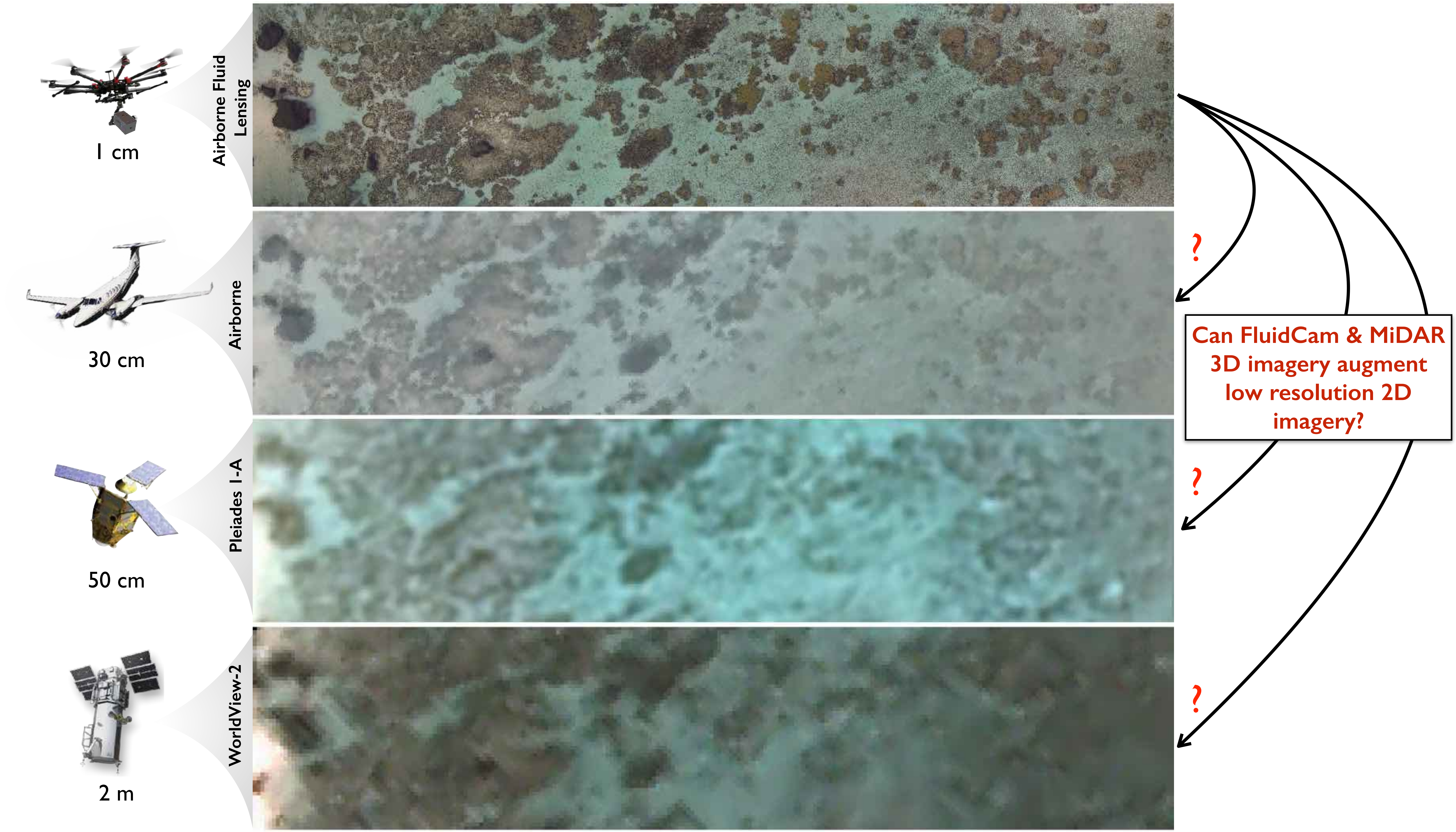
SILICON VALLEY
AMES RESEARCH CENTER
Student Participation

OCEANS & CLIMATE CHANGE

FluidCam 1&2 CUBESATS

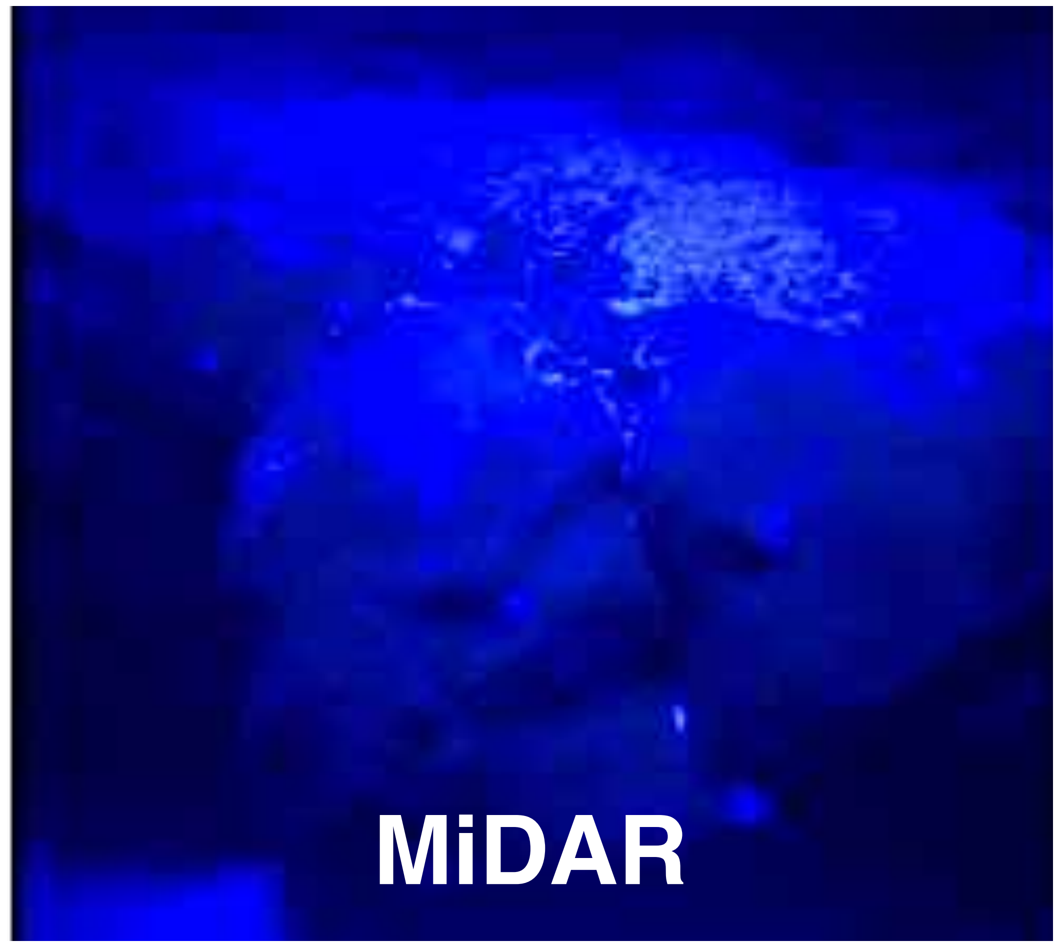


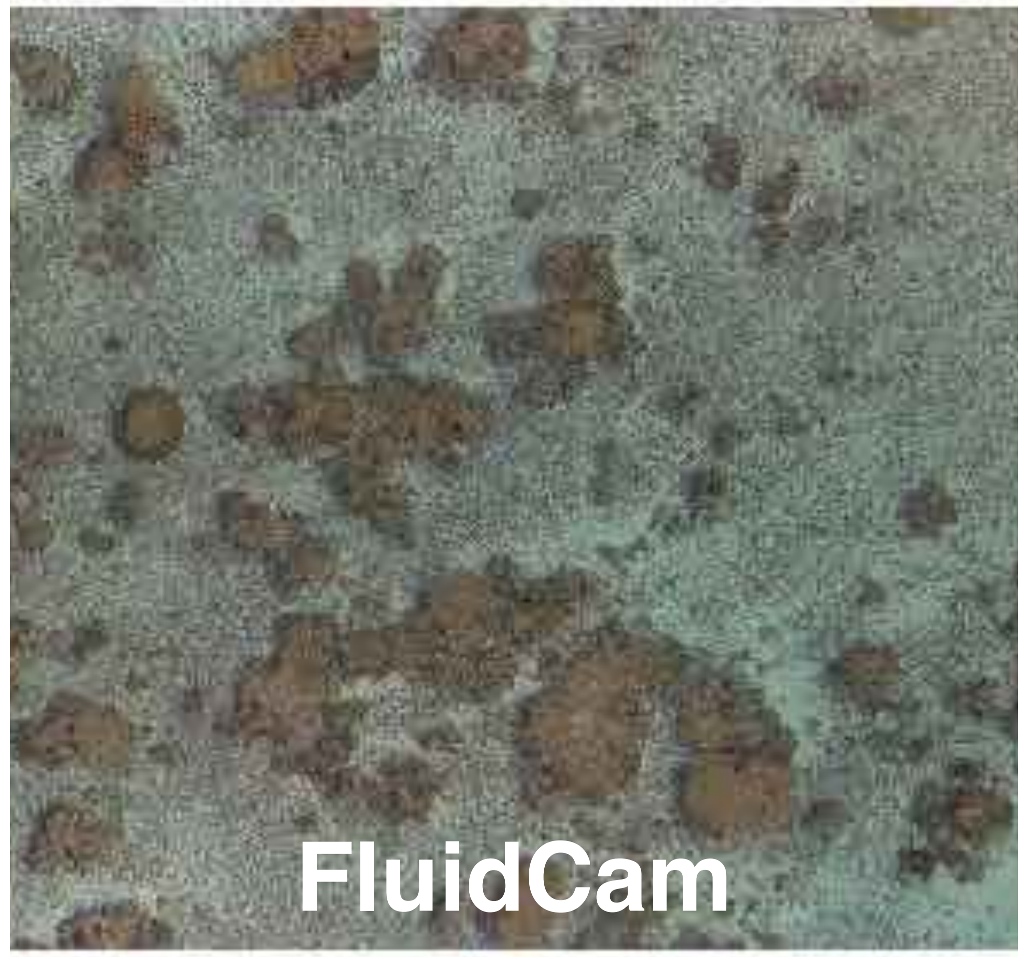

FluidCam
NASA AMES
LABORATORY FOR ADVANCED SENSING

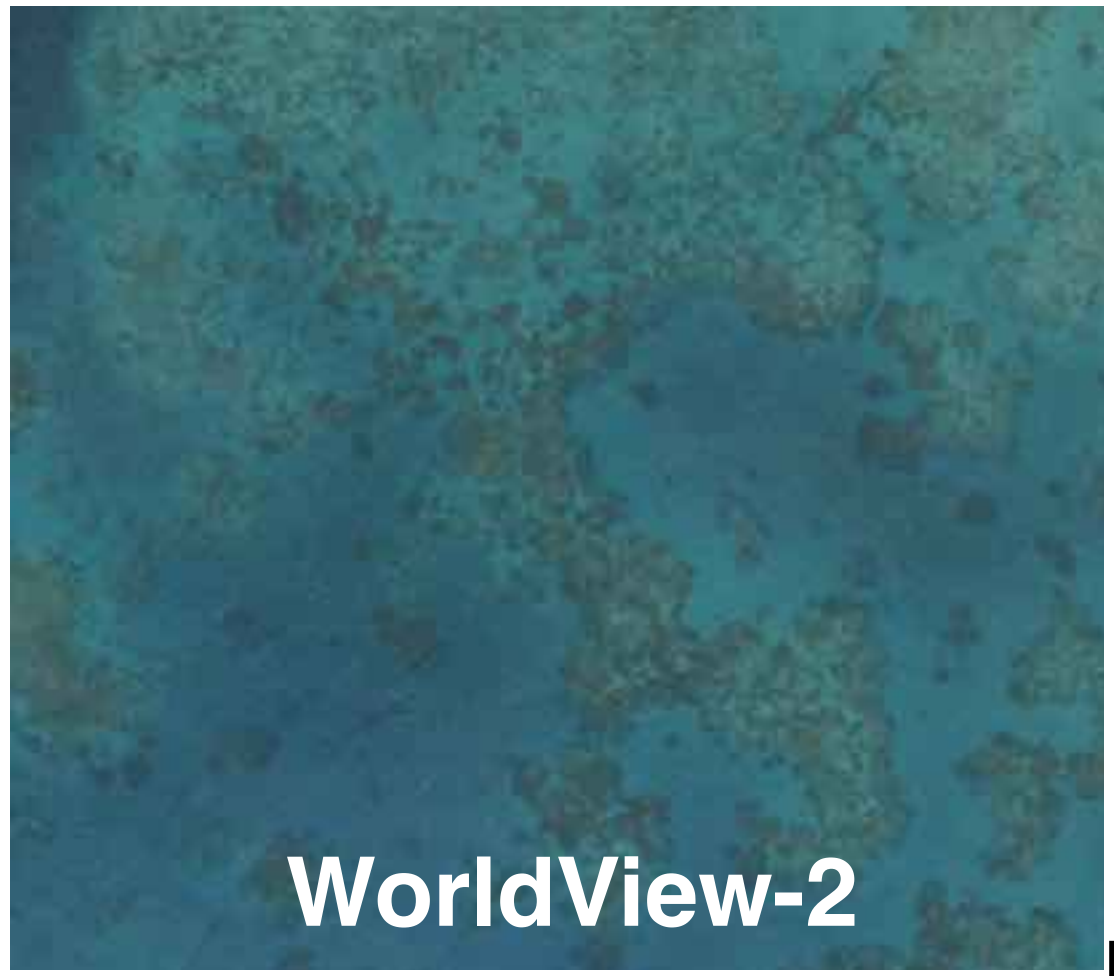

1 cm
Airborne Fluid Lensing
30 cm
Airborne
50 cm
Pleiades 1-A
2 m
WorldView-2
?
Can FluidCam & MiDAR 3D imagery augment low resolution 2D imagery?
?
?

NEMO-NET DATA SOURCES

Sensor	Effective Spatial Resolution	3D	Spectral Bands	Locations
Underwater AUV	0.1 - 5 cm	YES	3	Australia, Great Barrier Reef, Pacific
FluidCam & MiDAR (NASA)	**0.1 - 1 cm**	**YES**	**3-8**	American Samoa, Guam, Western Australia, Puerto Rico, Indo-Pacfifc
QuickBird (USGS)	0.65 m	NO	4	US Territories
WorldView-2/3	**0.5 - 3 m**	**NO**	**8**	Global
LandSat (USGS)	**30 m**	**NO**	**11**	Global







CLASSIFICATION IS HARD & CNNS NEED TRAINING DATA

Chihuahua vs. Blueberry Muffin

Bagel vs. Sleeping Dog

Mop vs. Hungarian Sheepdog


NASA NEMO NET
LAUNCHED EARTH DAY 2020 - NEMONET.INFO

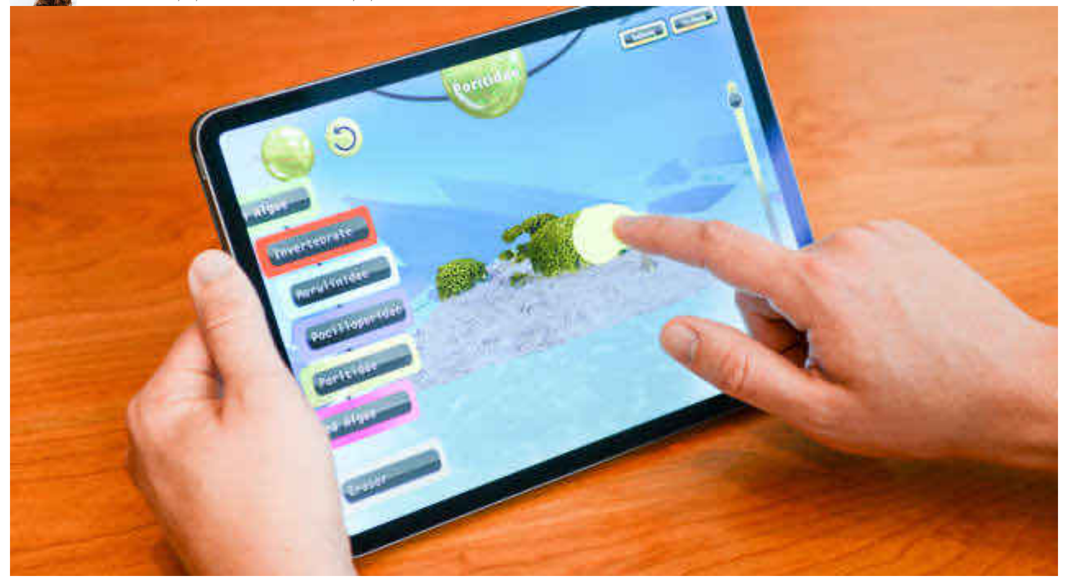
NEMO-NET REACHES >300M IN <MONTH

THRILLIST

ENTERTAINMENT

NASA Wants You to Help Save the Ocean's Corals by Playing This New Game

By AMBER SUTHERLAND-NAMAKO By AMBER SUTHERLAND-NAMAKO
Published On 04/09/2020 Published On 04/09/2020



NASA/Ames Research Center/Ved Chirayath

NASA's been using its sophisticated instruments and supercomputers to peek beneath the surface of the ocean to examine at-risk ecosystems for quite some time, but there's only so much it can do on its own. Cue: you.

Go ahead and start callin' yourself an honorary astronaut, because the National Aeronautics and Space Administration needs you to help save our precious coral reefs by playing a computer game in your underpants. To begin your next act as an amateur scientist, download the Apple-only NeMO-Net game from the App Store and dive into a virtual ocean tour. Traveling in your trusty vessel, the Nautilus, you'll explore NASA's own 3D images to find and categorize types of coral and ocean life.

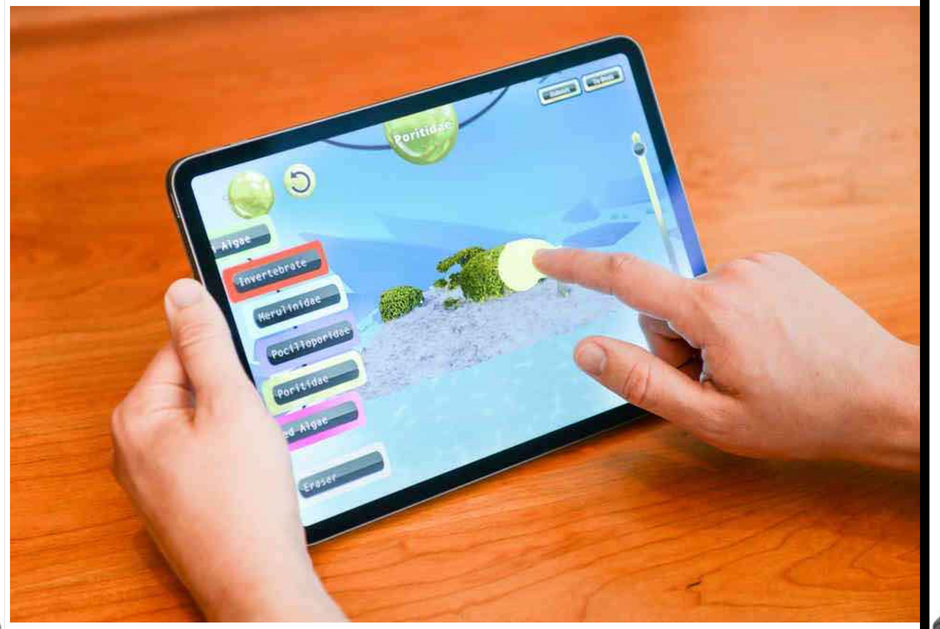
NeMO-Net's an eco game, and a numbers game -- the more people who play, the more insight NASA can glean.

NASA prosí o pomoc: Hraním hry NeMO-Net pomůžete vědcům učit superpočítač

před 8 hodinami Žádné komentáře

Hraním her na mobilu a tabletu většinou mnoho užitečného nedosáhneme. To se však nedá říct o aplikaci NeMO-Net od agentury NASA. Chytrý počin pomáhá naučit superpočítač rozpoznávat korály a analyzovat 3D snímky ze satelitů. Pomoci může každý, je to příjemná oddychová zábava.

وكالة ناسا تُصدر لعبة فيديو ستسمح لك بالمساعدة في رسم خريط لشعاب المحيط المرجانية



美国NASA希望你能帮助绘制全球珊瑚地图

2020年04月11日 00:10 362 次阅读 稿源：cnBeta.COM 0 条评论

美国NASA呼吁公民科学家通过玩一个虚拟潜水游戏来帮助识别和分类世界上的珊瑚，让专家们能够更好地了解它们的进化，以及如何在未来保护它们。用户需要的是下载NASA NeMO-Net游戏，它目前有iOS和iPadOS 版本（macOS和Android版本即将推出）。

这款游戏将带你在海洋中进行一系列的虚拟潜水，你的任务是识别你所遇到的珊瑚。计算机生成的水下环境是基于美国宇航局位于加州的艾姆斯研究中心在过去几年中收集到的数据。那里的团队一直在使用流体透镜相机绘制出比以往更详细的海洋地图。

这些相机最初是为了让地面上的天文学家能看到不受大气层扭曲的恒星而开发的，但也可以避免水的扭曲来绘制海底地图。然而，尽管这些相机再先进，但它们并不能揭示出坐在海浪下的珊瑚的全部图片，通过巡视和识别你在游戏中看到的珊瑚类型，以及它们的确切位置，你可以帮助NASA收集相关数据。

NASA表示任何人，甚至是一年级的小学生，都可以通过玩这个游戏，对这些数据进行分类，帮助我们绘制出全球珊瑚地图。这个游戏也很有教育意义，教用户识别世界海洋中珊瑚的种类。用户提交的所有资料都会被艾姆斯研究中心的Pleiades超级计算机处理，训练它如何根据原始数据识别不同类型的珊瑚。与其他神经网络一样，随着训练的深入，它的识别效率应该会随着时间的推移而不断提高，因此，即使是质量较低的数据，它最终也能独立识别出珊瑚类型。注册使用NeMO-Net的人越多，系统就会变得越聪明。

NASA построит карты коралловых рифов в океане с помощью игры NeMO-Net

13 апреля 2020, 23:45

Проблема выявления и классификации коралловых рифов давно значится в числе приоритетных задач для человечества. Сейчас изучение поселений коралловых рифов, их динамики ведется с использованием снимков, полученных из космоса. Но на этом пути есть серьезная проблема.

Ее разъяснение недавно появилось со стороны NASA. Представители космического агентства разъяснили, что у них имеются сейчас все возможности и **технологии** для получения трехмерных изображений коралловых рифов в океане. В распоряжении исследовательского центра NASA в Калифорнии есть необходимые приборы, которые позволяют наблюдать за изменениями, происходящими с коралловыми рифами. Наблюдение возможно, даже несмотря на оптические искажения на снимках, вызываемые волнением водной поверхности. В результате NASA получает трехмерные изображения морского дна мирового океана, на котором нанесены изображения коралловых рифов и колонии морских растений.

Но существует проблема. Она состоит в том, что для их идентификации и классификации рифов необходимо использовать труд человека. Несмотря на простоту задачи распознавания рифов, достоверно выполнить эту задачу может только человек.

Поэтому в NASA объявили о выпуске **игры** под названием «NeMO-Net». Геймплей связан с захватом игроком колоний коралловых рифов. Занимаясь этим, геймеры показывают, как идентифицировать и классифицировать захваченные ими рифы. По мнению представителей NASA, игра доступна для понимания даже самым маленьким. Они смогут играть и помогать ученым в процессе сортировки данных об океане.

Название игры взято по имени нейронной, мультимодальной сети, на базе которой она построена. В процессе игры участники помогает обучить суперкомпьютер NeMO-NET. Он учится распознавать коралловые рифы из тех элементов, которые ему указывают игроки. Со временем суперкомпьютер узнает, что представляет собой каждый элемент кораллового рифа. В результате формируется их классификация. Ее создают вручную игроки, прокладывая себе путь в процессе освоения территорий.

El Sol de México

Con videojuego NeMO-Net, NASA buscará corales

Los arrecifes de coral del fondo marino de las regiones subtropicales y tropicales del planeta albergan algunos de los ecosistemas más diversos y complejos del planeta

Foto: Reuters

benchmark.pl » Gry » Zagraj w grę i pomóż NASA ocalić rafy koralowe

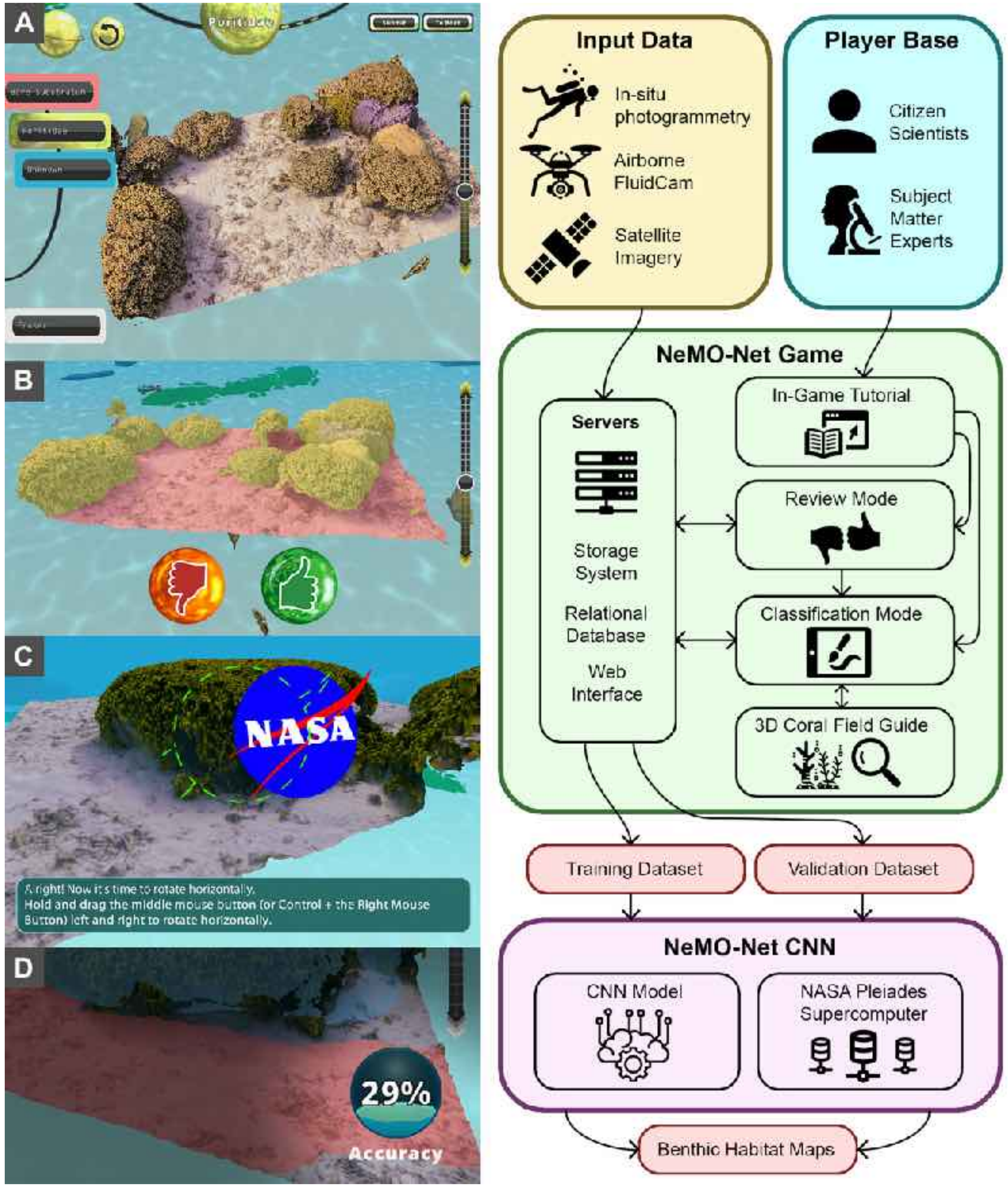
GRY

Zagraj w grę i pomóż NASA ocalić rafy koralowe

z dnia 12-04-2020

FIELD CAMPAIGNS

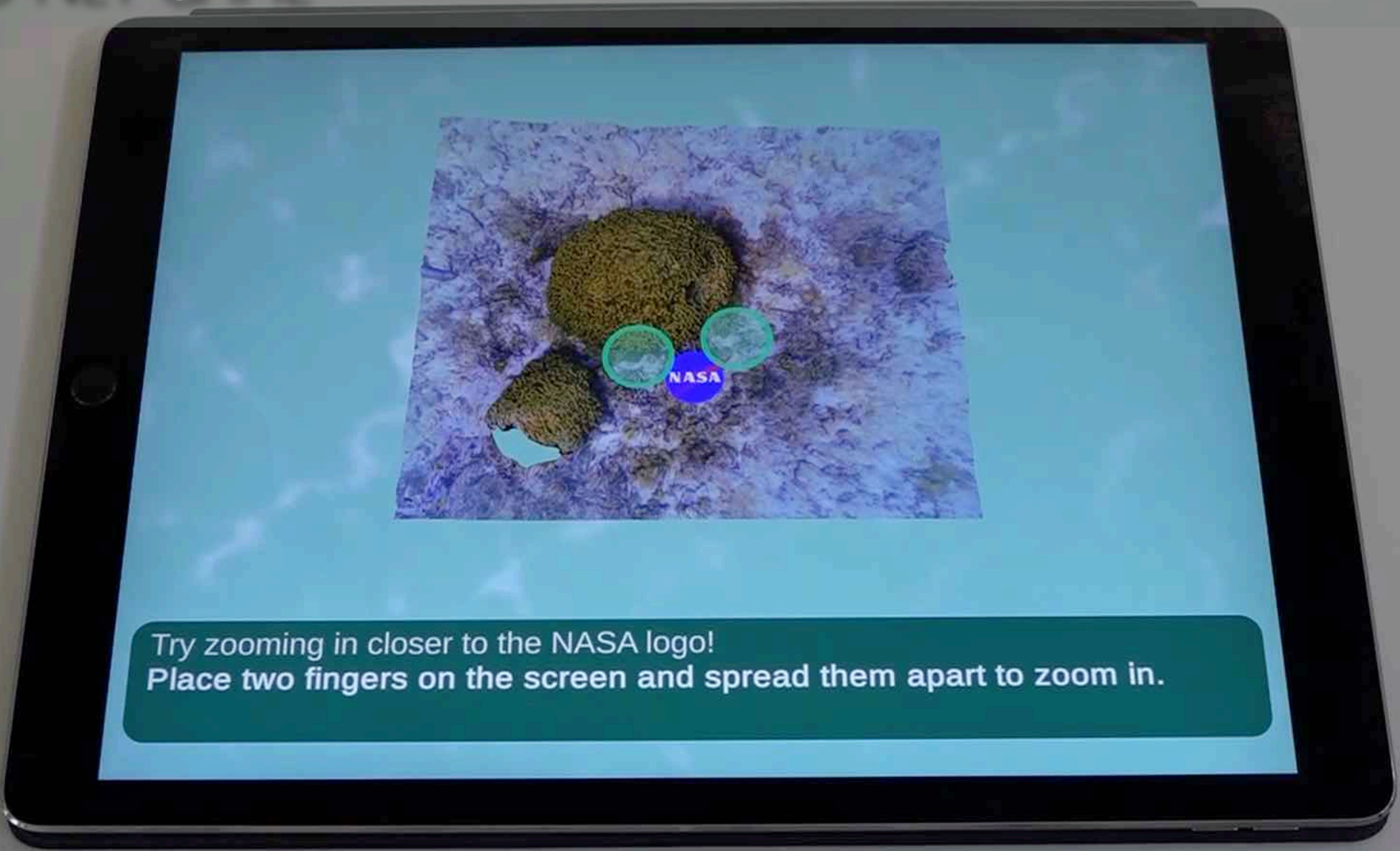
NEMO-NET GAME OVERVIEW



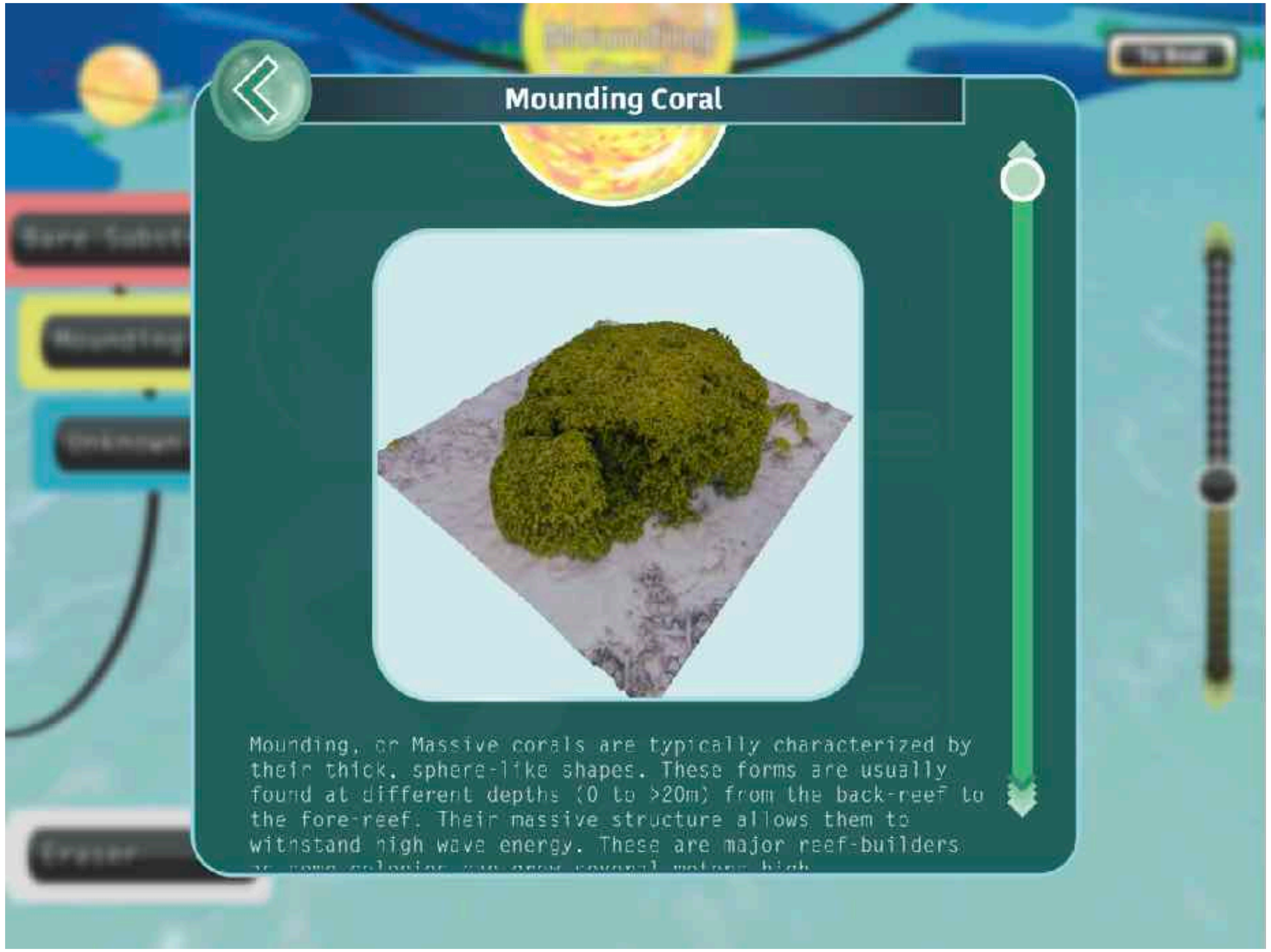
- Now available on Windows, Mac, Android and iOS!
- New Frontiers paper in review!
- New language support coming soon!

NEMO-NET GAME

NEMO-NET GAME

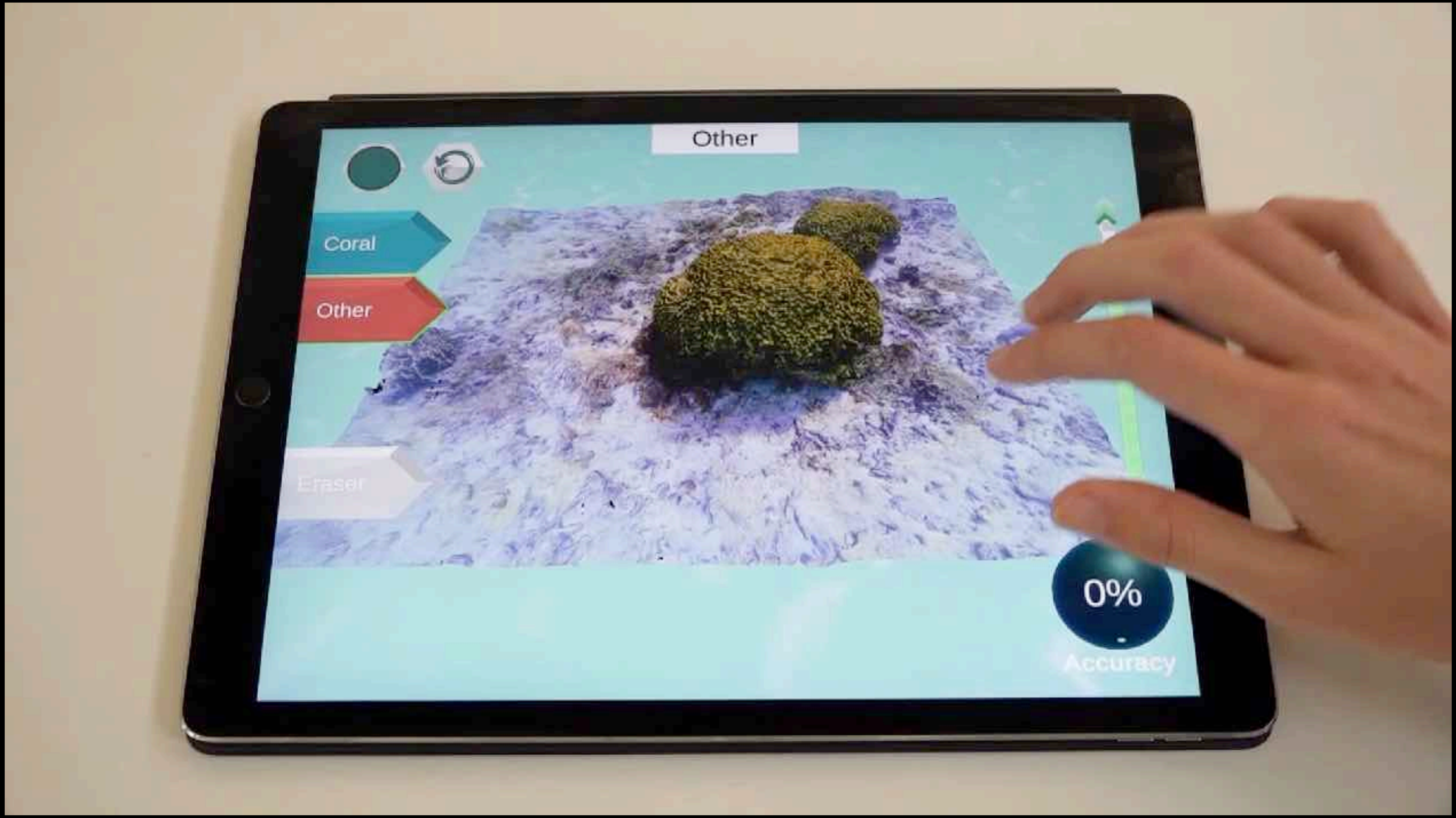


NEMO-NET TRAINING & LEVELS

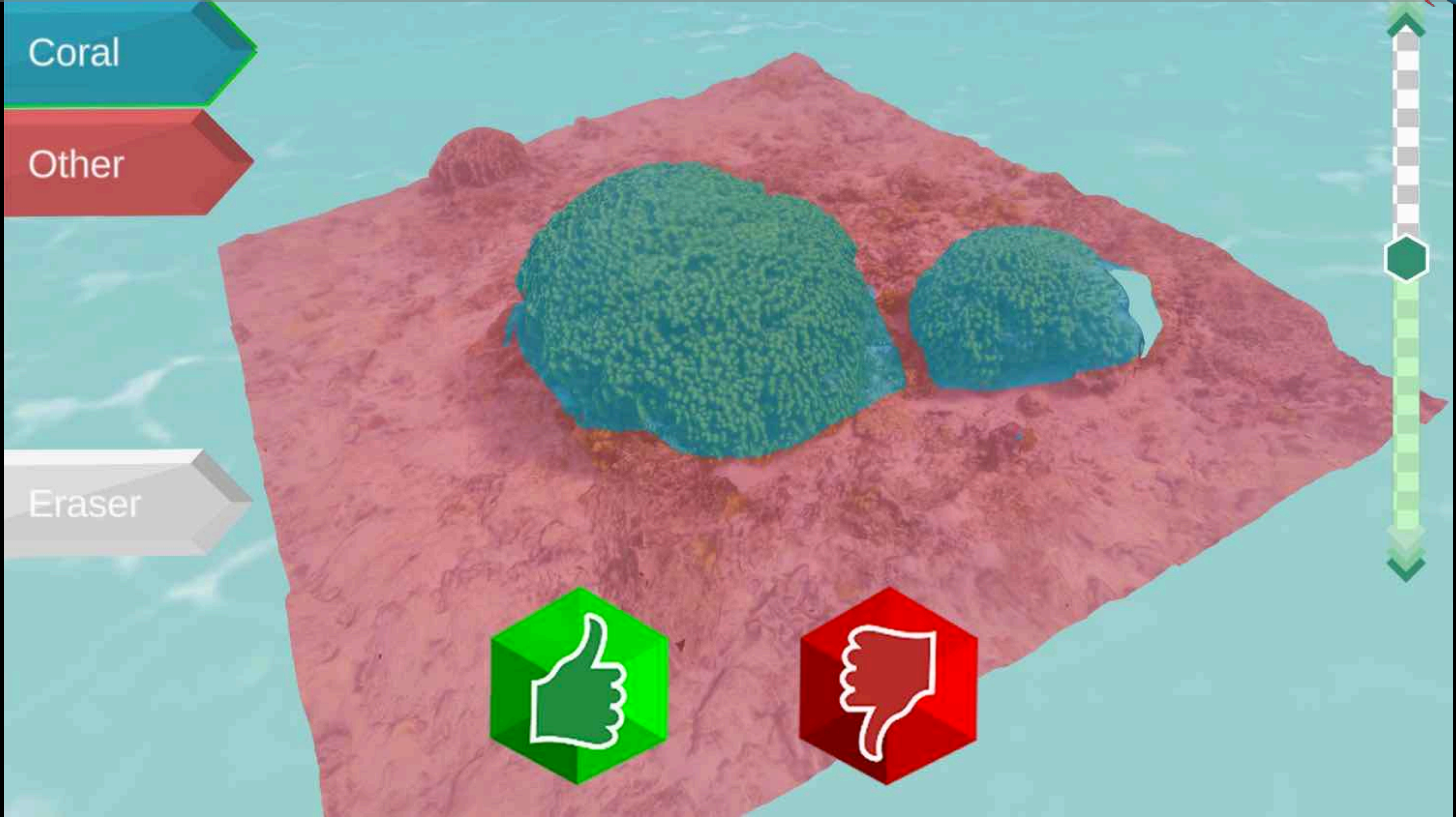


NEMO-NET GAME

Players are required to classify coral at a minimum accuracy before sending data to the neural network



ACTIVE LEARNING

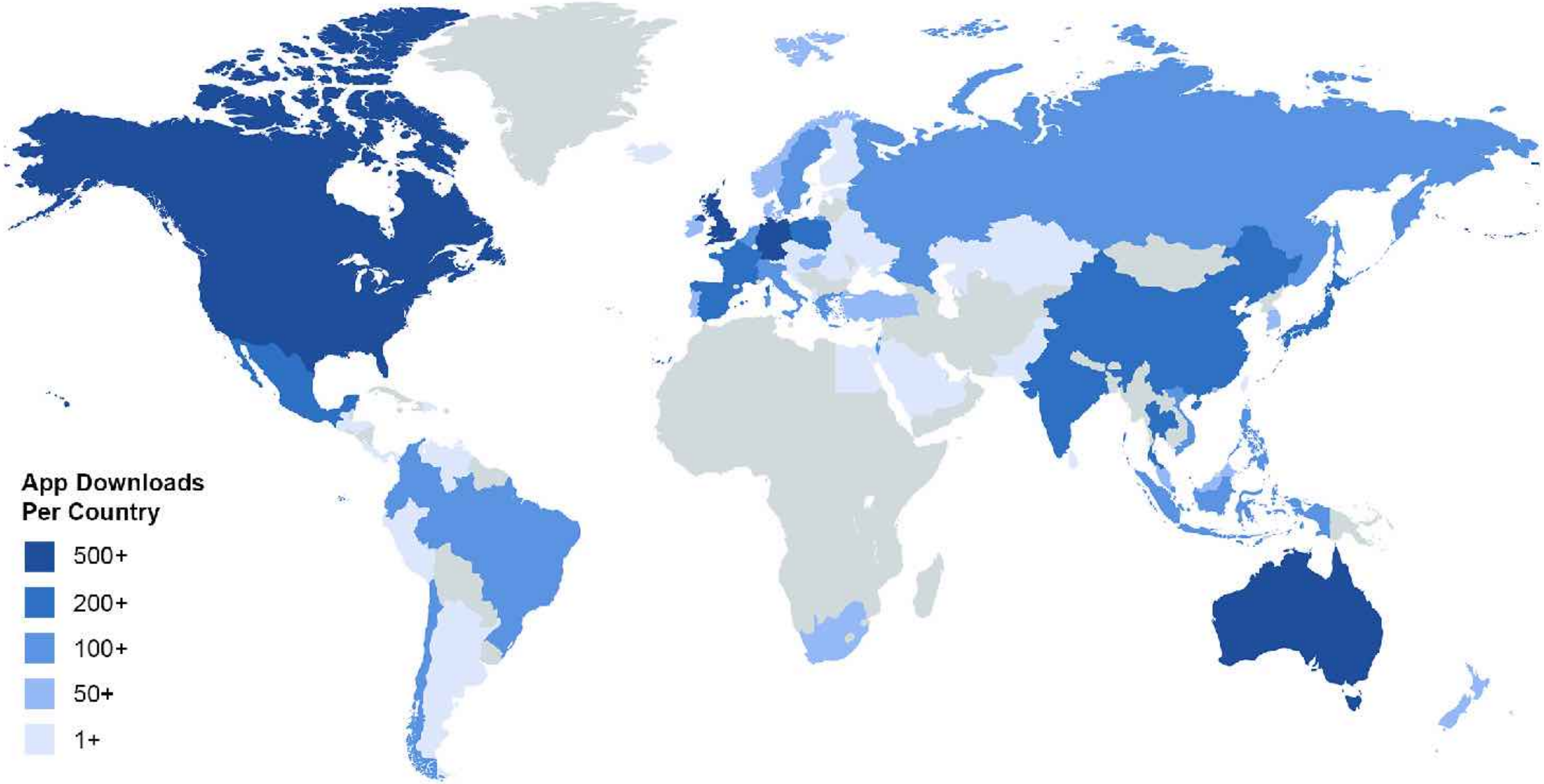


GAME STATISTICS

- 78,646 classifications
- 64,728 classification ratings
- 45,800 unique downloads

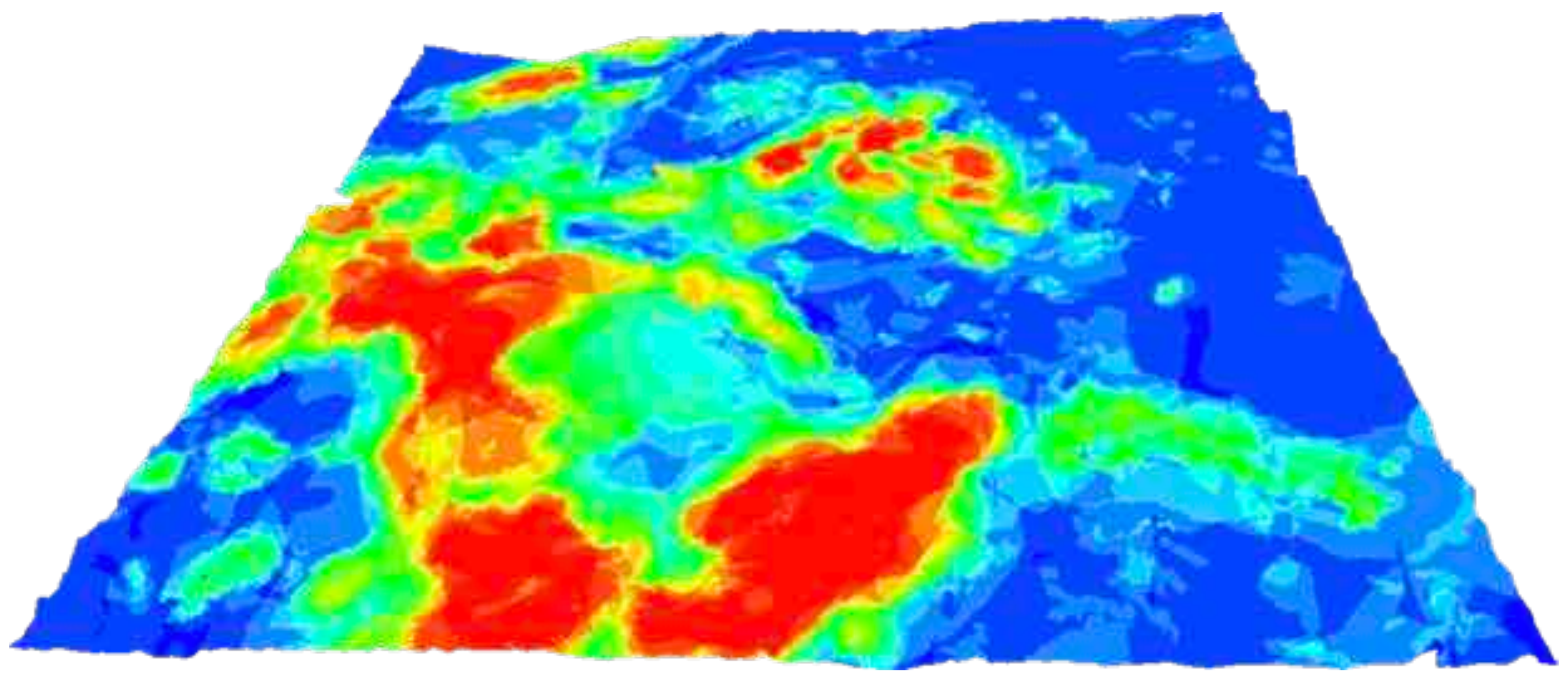
- 4.9 star rating on iOS
- 4.3 star rating on Mac
- 4 star rating on Android

- Reached #3 for family apps on Mac App Store
- Reached #48 for casual apps on iOS App Store

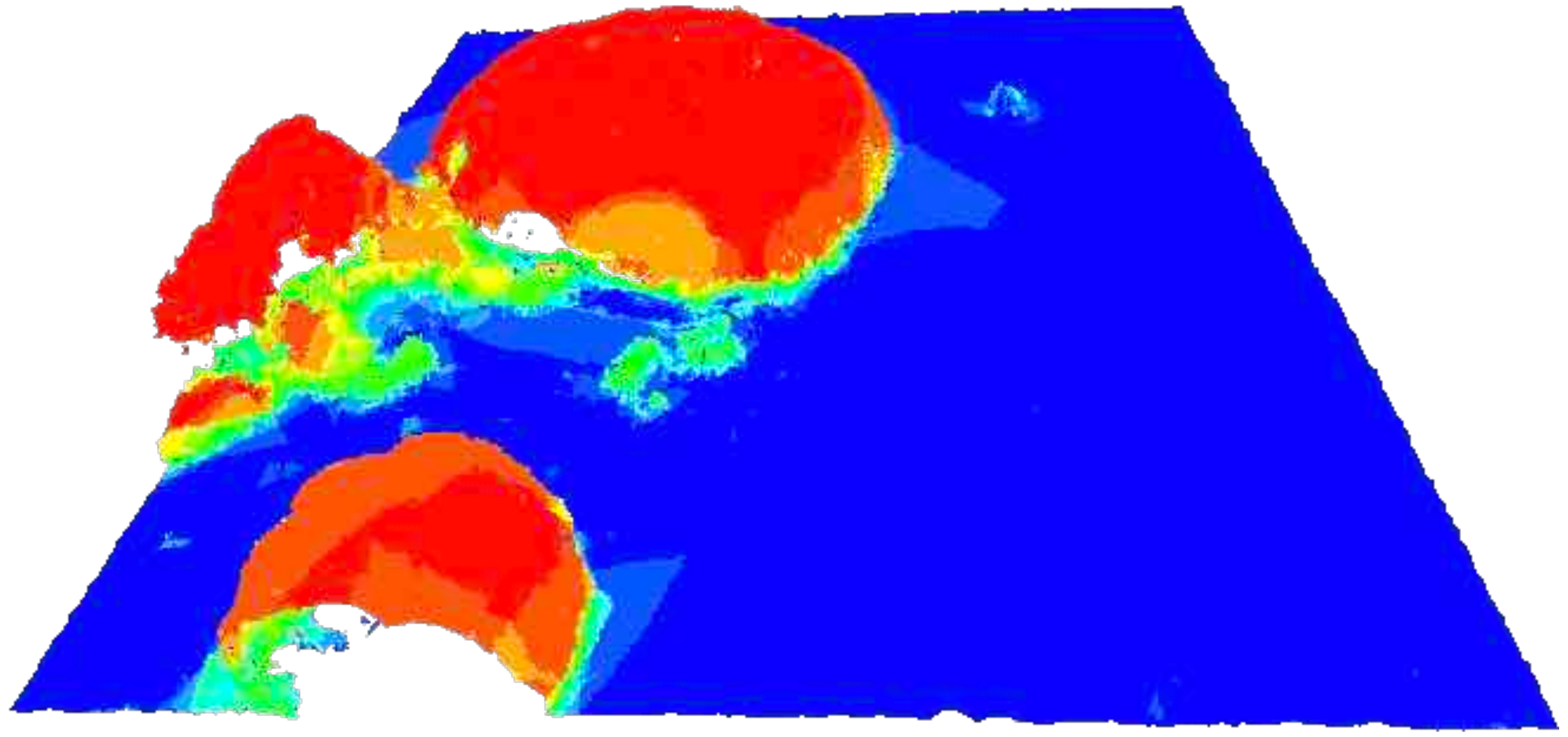


CLASSIFICATION CONVERGENCE

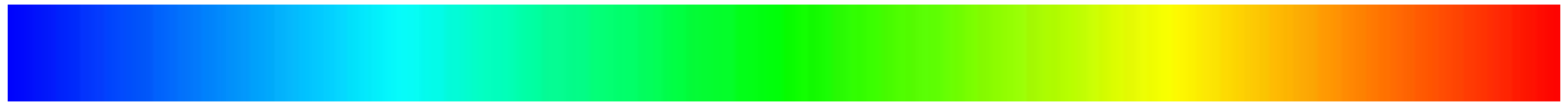
Heat Map of Algae



Heat Map of Coral

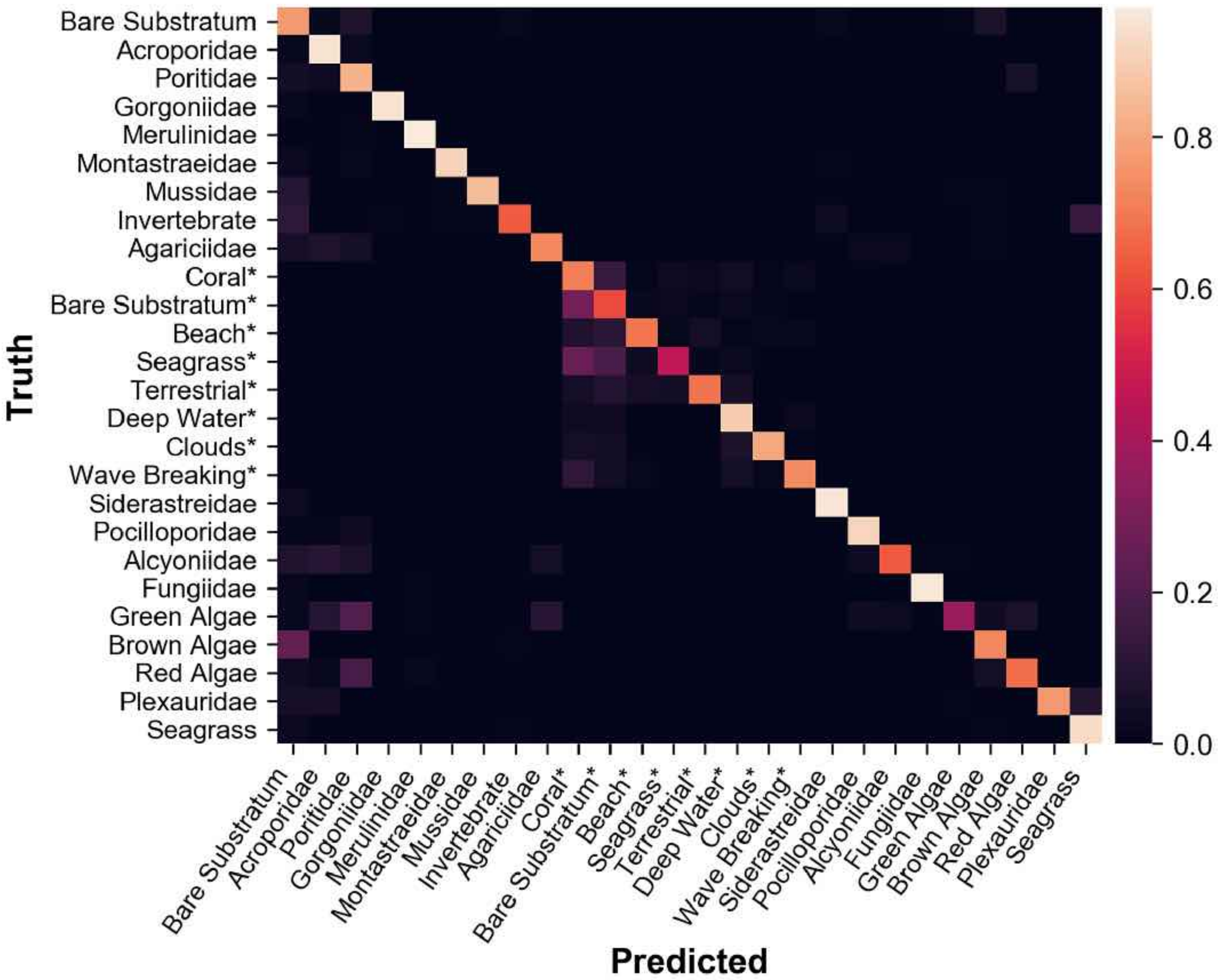


0% of Users Classified as Algae / Coral



100% of Users Classified as Algae / Coral

CLASSIFICATION ACCURACY



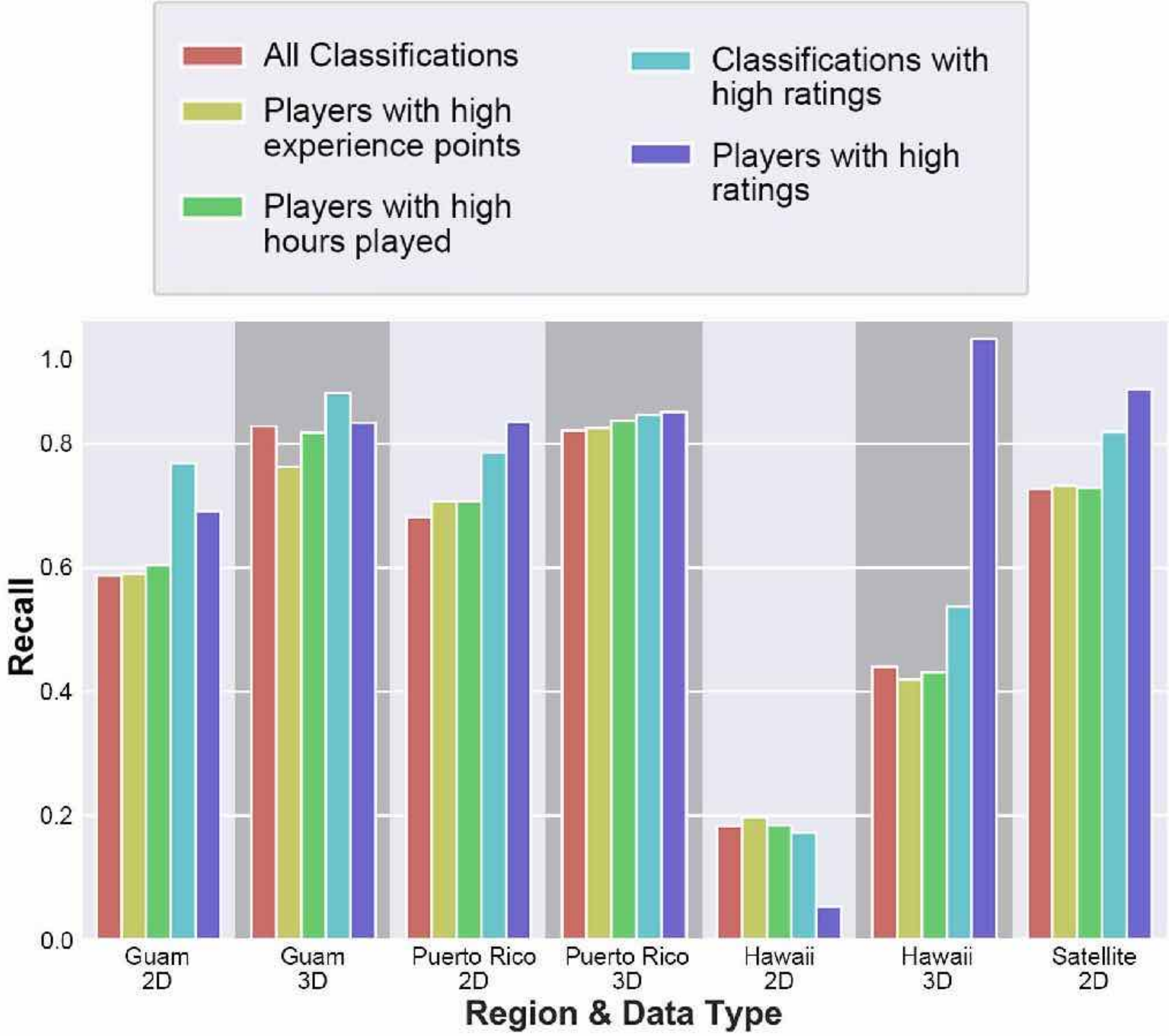
Calculated by comparing player classifications on pixel-by-pixel basis to a representative truth dataset of 134 expert 3D classifications.

Average *recall* (accuracy) over all class labels is ~78%

Average *precision* over all class labels is ~57%.

Class labels with an asterisk indicate that the class label is used purely in satellite classifications.

CLASSIFICATION FILTERS



Classification ratings is an effective filter!

Each of NeMO-Net's regions has a higher 3D recall than 2D recall.

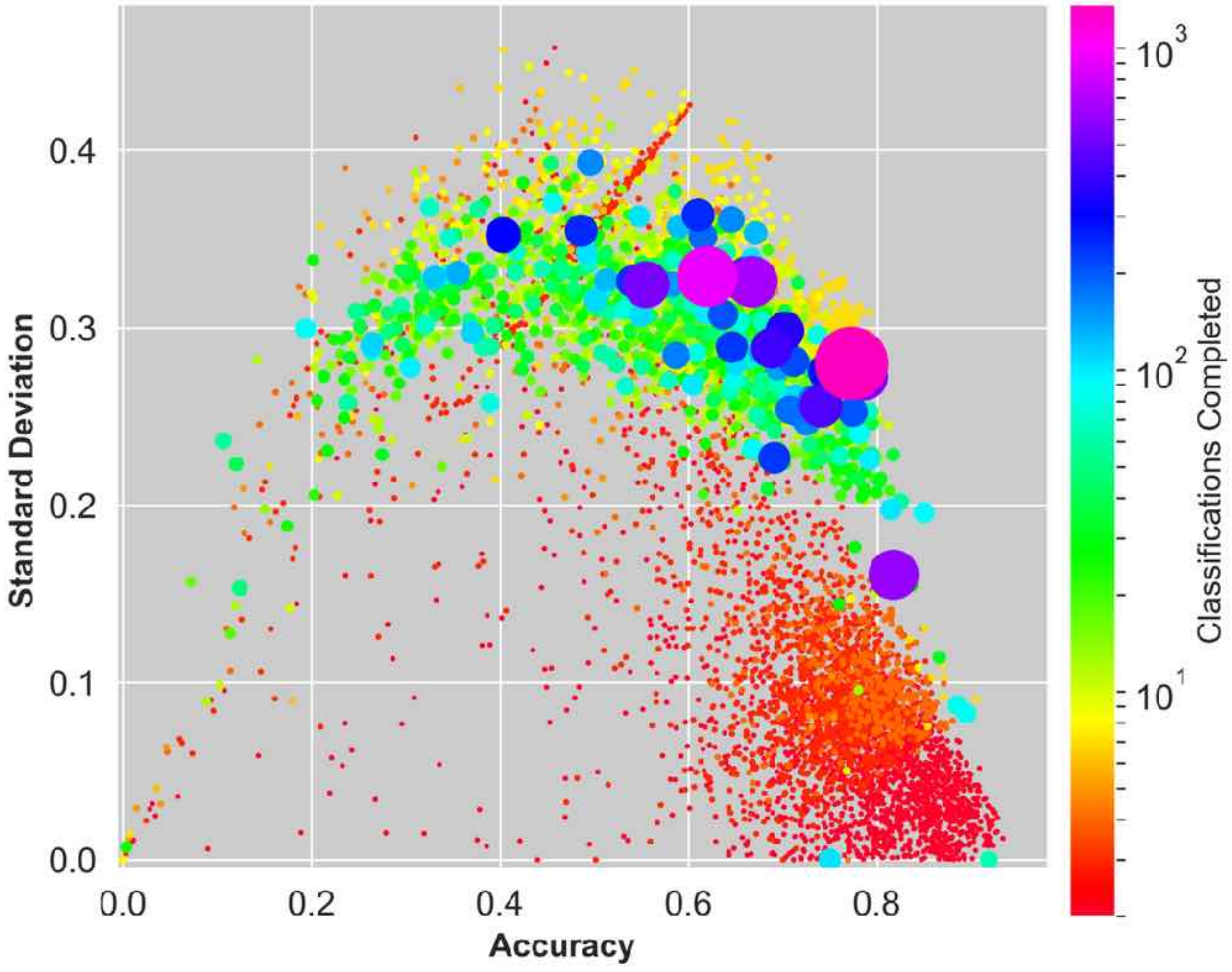
- Guam (all classifications) has a 41.14% increase

- Puerto Rico (all classifications) has a 20.26% increase

- Hawai'i (all classifications) has a 140.45% increase

- High rated Hawai'i players have a 1739.9% increase

USER STATISTICS

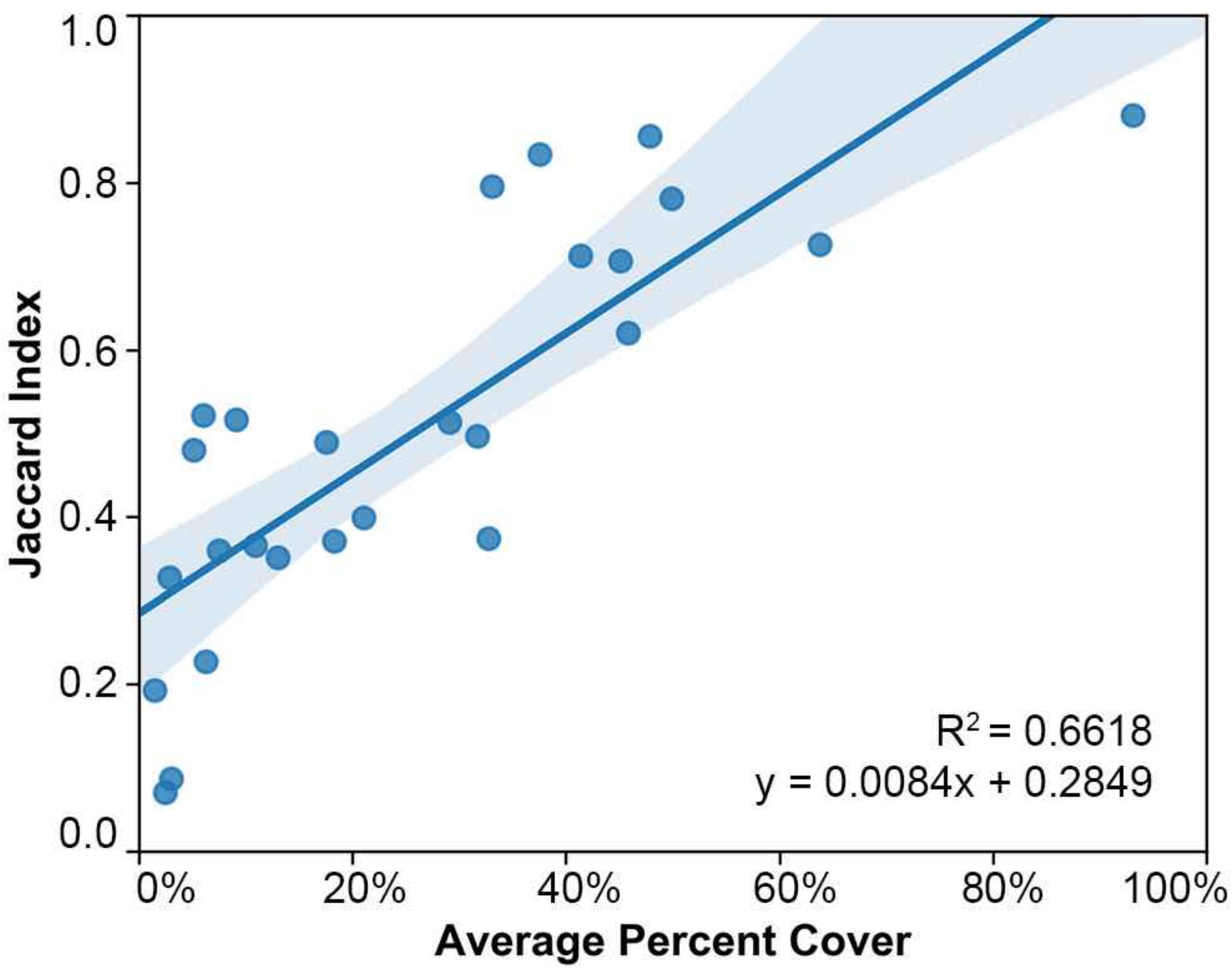


Each datapoint represents an individual player, the size and color of the datapoint corresponds to the classifications completed.

Standard deviation can be used to estimate player consistency.

A small group of players have submitted several classifications, have high average accuracy, and a relatively low standard deviation.

AVERAGE PERCENT COVER



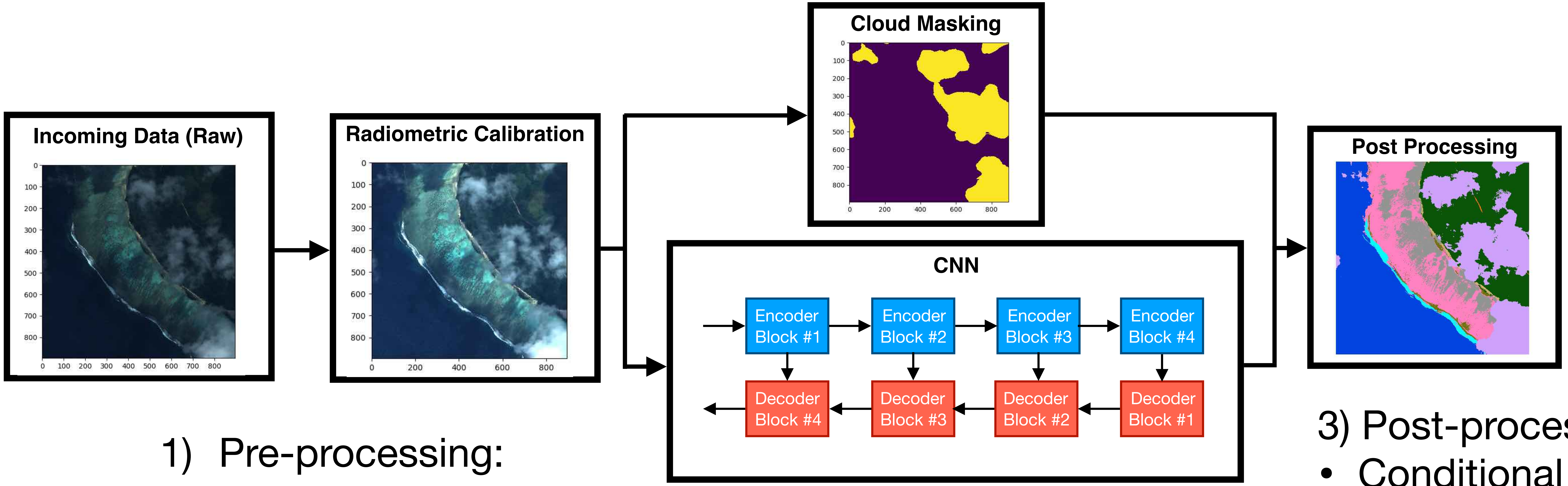
Average Percent Cover is what percent of a transect a class label covers on average. Jaccard Index is a measure of overall classification quality. Each datapoint represents an individual class label.

Class labels with high average percent cover have higher classification quality.

Class labels with high average percent cover improve in accuracy as classification ratings improve.

NEMO-NET MACHINE LEARNING OVERVIEW

- The NeMO-Net Machine Learning algorithm is split into 3 sections:



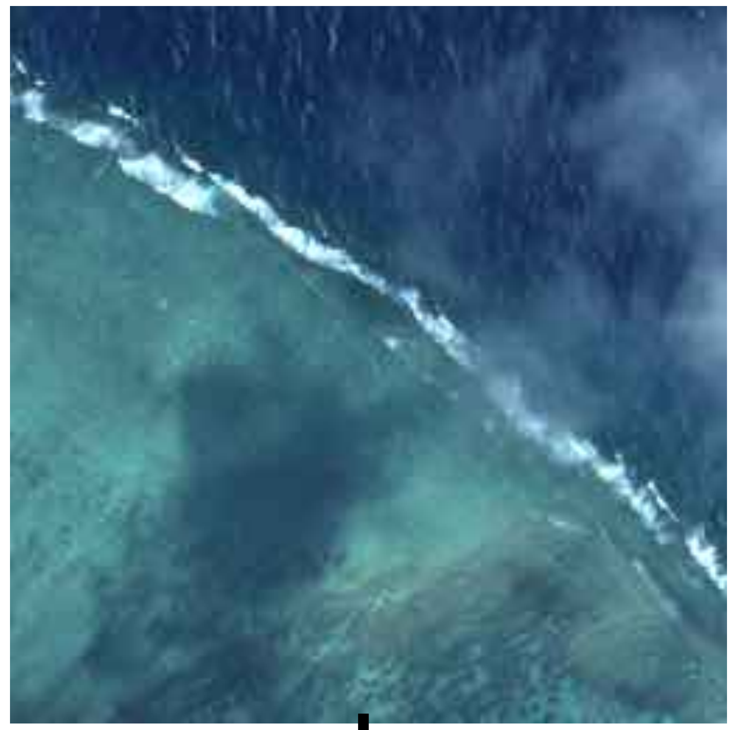
1) Pre-processing:
- File and data preparation
- Augmentation
- Calibration

2) CNN
- Image segmentation and classification

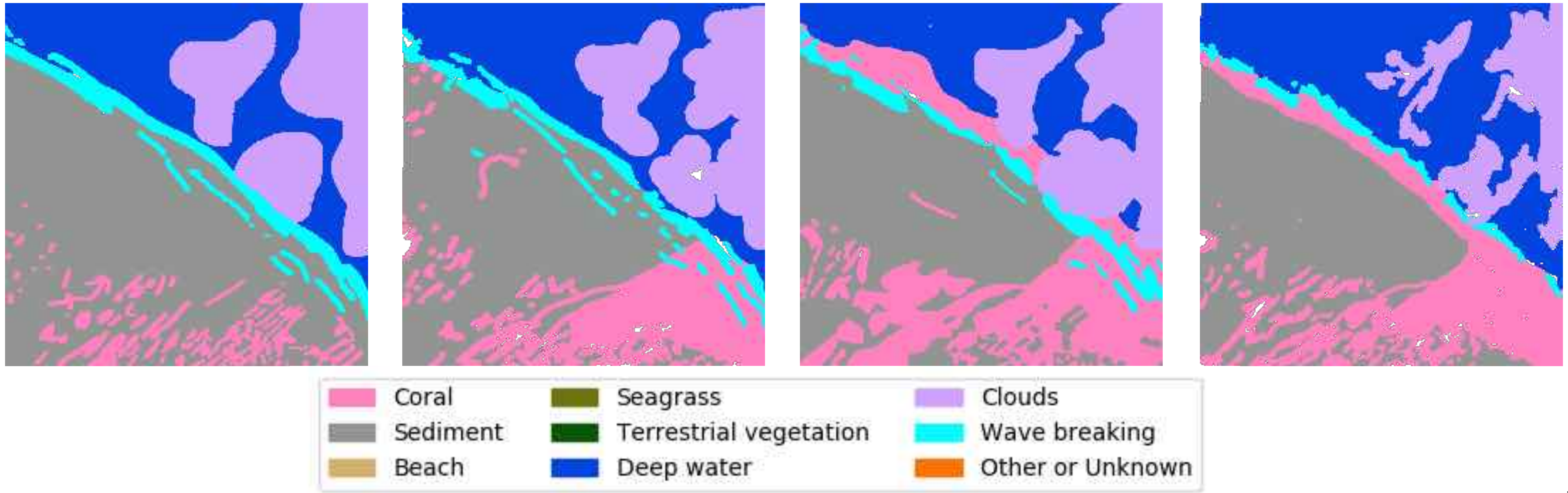
3) Post-processing
- Conditional random field (CRF)
- K-Nearest neighbors (KNN)

PRE-PROCESSING

RGB Imagery



User Classifications



Digital Numbers (DN) to Top of Atmosphere (TOA) radiance

Solar Angle Correction

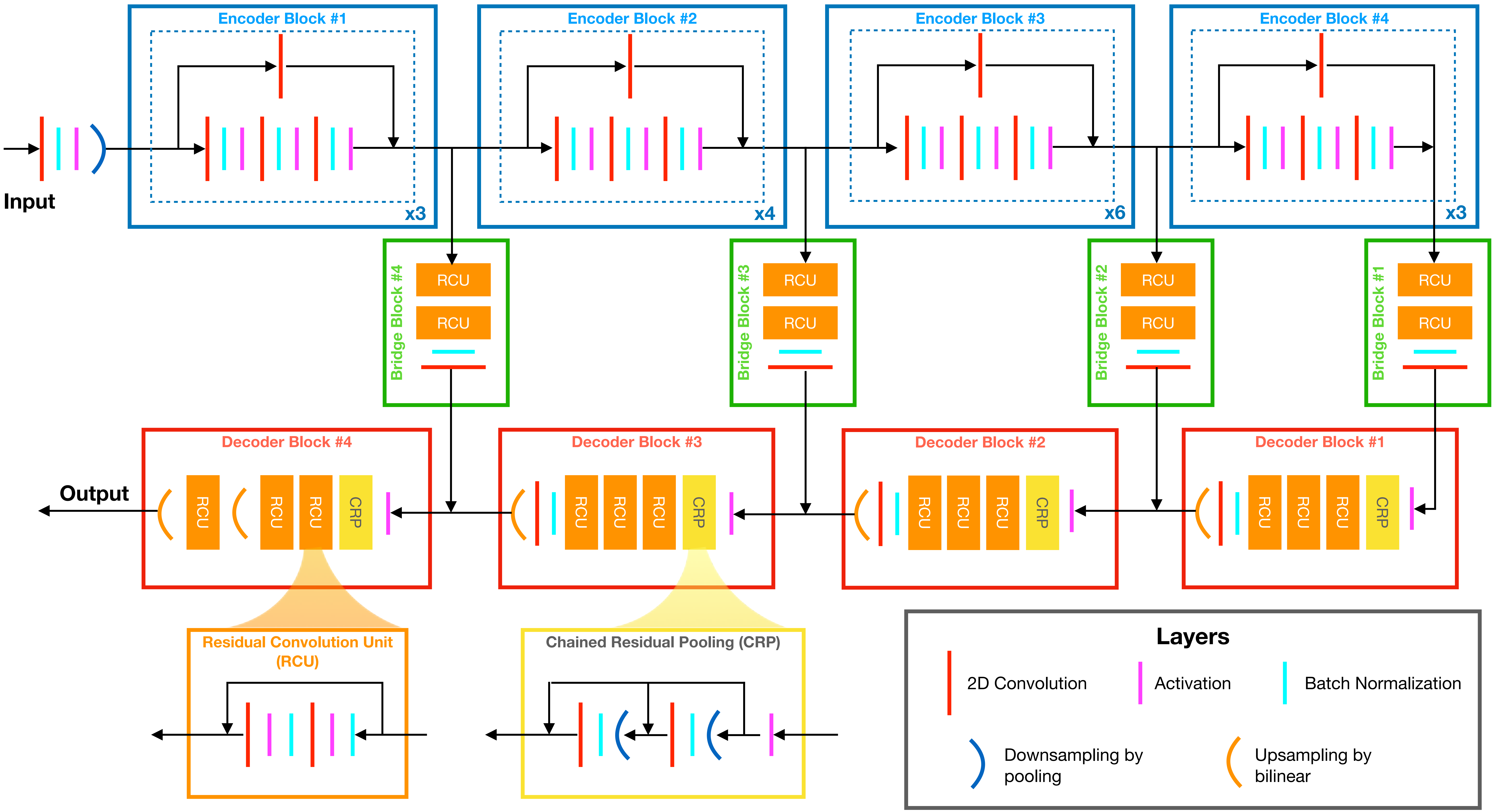
Data Augmentation (Spectral randomization, image modification)

Trusted user screening (known scientists and users)

To CNN

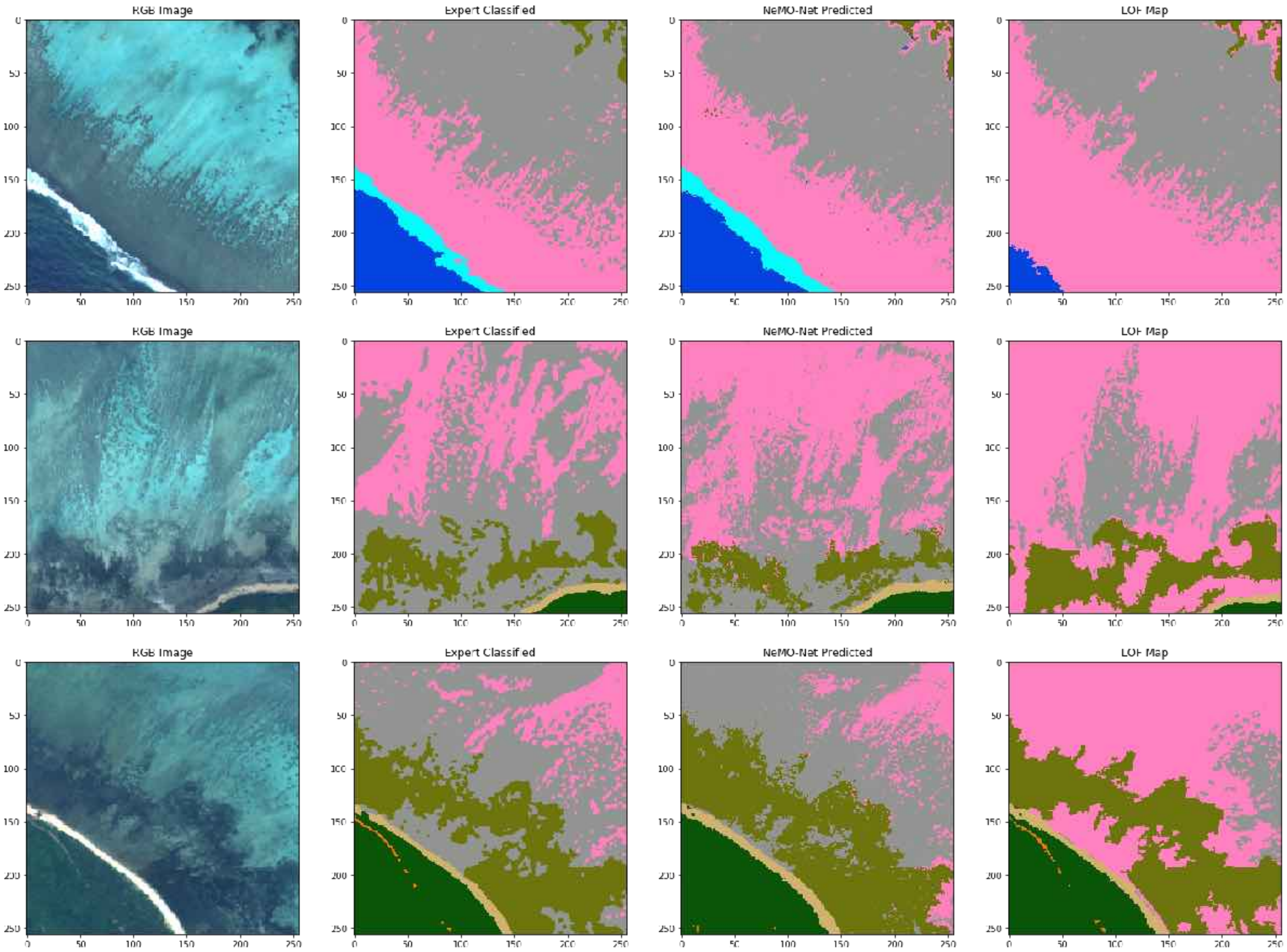
Data Sources:
- WorldView-2 (Living Oceans Foundation)
- Planet satellites

CNN



Input
Encoder Block #1
x3
Encoder Block #2
x4
Encoder Block #3
x6
Encoder Block #4
x3
Bridge Block #4
Bridge Block #3
Bridge Block #2
Bridge Block #1
RCU
RCU
Decoder Block #4
Decoder Block #3
Decoder Block #2
Decoder Block #1
CRP
Output
Residual Convolution Unit (RCU)
Chained Residual Pooling (CRP)
Layers
2D Convolution
Activation
Batch Normalization
Downsampling by pooling
Upsampling by bilinear

PERFORMANCE METRICS: WV-2

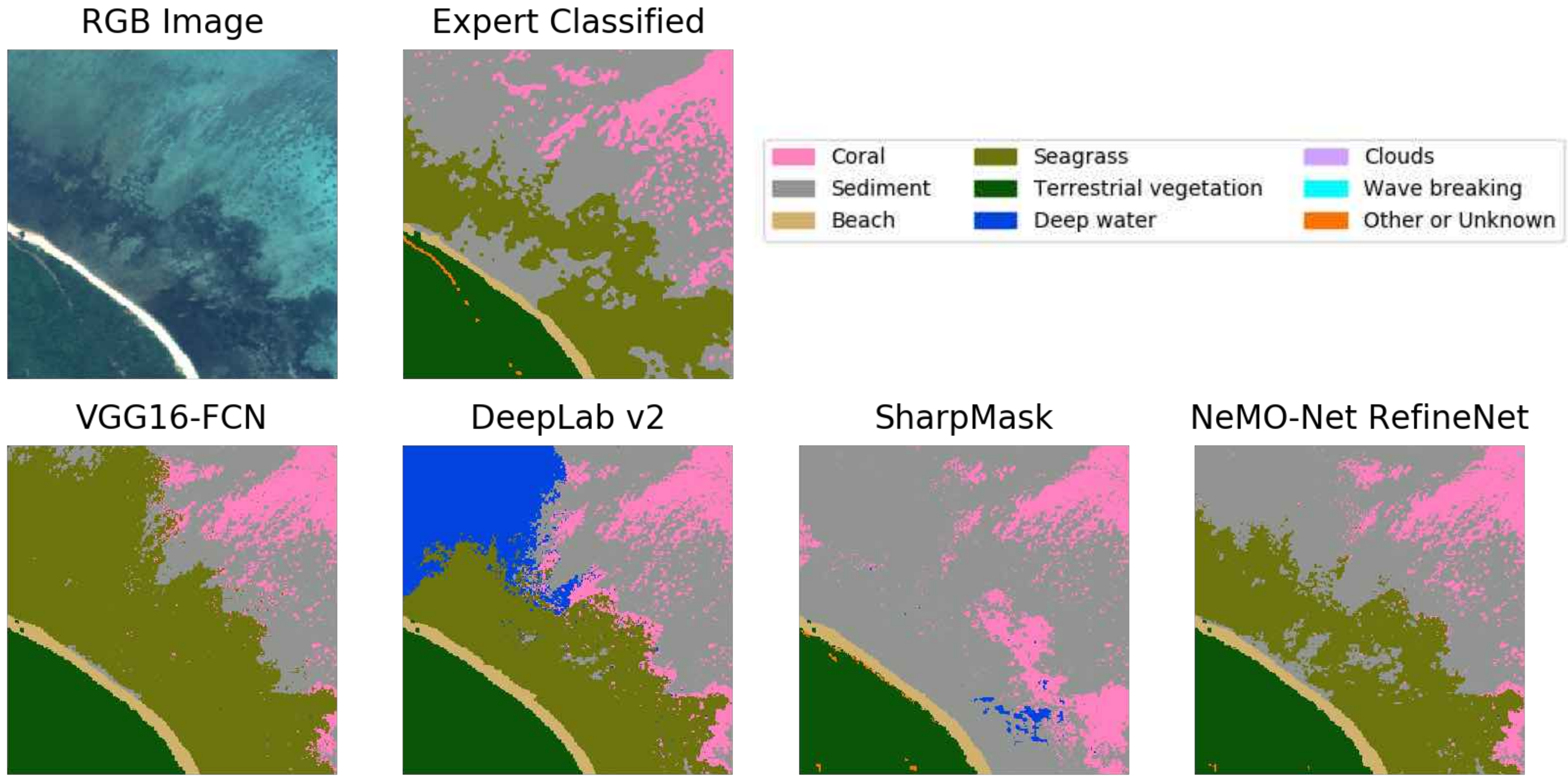


Confusion matrix, without normalization

True label \ Predicted label	Coral	Sediment	Beach	Seagrass	Terrestrial vegetation	Deep water	Wave breaking	Other or Unknown
Coral	84514	12337	0	73	0	15	626	0
Sediment	23193	129335	456	7113	6	5	0	0
Beach	0	2	2345	0	68	0	0	0
Seagrass	3723	1976	179	32256	2	56	0	0
Terrestrial vegetation	0	1	179	0	11159	0	0	16
Deep water	421	86	0	2	0	9666	442	0
Wave breaking	103	164	0	0	0	6	6991	0
Other or Unknown	0	0	0	0	161	0	0	3

Total Accuracy: 84.3%

RESULTS



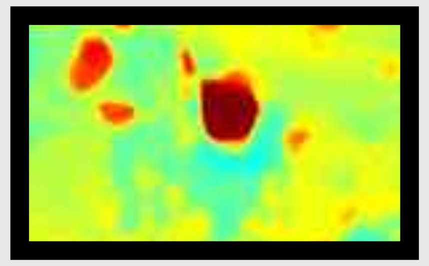
Comparisons against other common CNN architectures

Method	Accuracy	Mean Precision	Mean Recall	Frequency-weighted IOU
CNN-CRF only (all classes)				
VGG16-FCN	81.5%	**74.0%**	**77.5%**	69.2%
DeepLab	61.0%	62.0%	55.9%	43.6%
SharpMask	79.9%	68.8%	63.0%	66.6%
NeMO-Net (RefineNet) (Focal)	75.9%	73.4%	70.2%	61.5%
NeMO-Net (RefineNet) (Lovasz)	81.2%	70.4%	61.6%	69.6%
NeMO-Net (RefineNet) (CE)	**82.1%**	68.3%	67.8%	**69.9%**
Post-processing with KNN (all classes)				
VGG16-FCN	81.0%	73.7%	78.3%	68.8%
DeepLab	78.5%	67.4%	74.0%	65.4%
SharpMask	80.3%	68.3%	64.3%	67.5%
NeMO-Net (RefineNet) (Focal)	80.6%	73.8%	72.8%	67.7%
NeMO-Net (RefineNet) (Lovasz)	83.6%	69.3%	65.4%	72.4%
NeMO-Net (RefineNet) (CE) No CRF	83.5%	67.3%	69.6%	71.8%
NeMO-Net (RefineNet) (CE) with CRF	**84.3%**	**77.6%**	**79.5%**	**72.9%**
Post-processing with KNN-CRF (Coral, sediment, and seagrass only)				
VGG16-FCN	79.6%	76.2%	82.5%	66.7%
DeepLab	80.7%	77.8%	82.9%	67.9%
SharpMask	79.8%	78.2%	70.9%	66.2%
NeMO-Net (RefineNet)	**83.6%**	**82.6%**	**84.4%**	**72.0%**
KSLOF Ecognition Prediction	65.1%	68.9%	59.0%	48.6%

- Post-processing gets the accuracy to ~83.6%
- Weighted cross-entropy loss and Lovasz loss performed the best generally

MM-Scale Airborne Fluid Lensing

MM-Scale Airborne Fluid Lensing DEM



M-Scale Airborne & Satellite Data

VR & App-based Active Learning & Interactive Training through IUCN, Mission Blue, & Partners

Level 1 Data & Existing Training Data Analysis

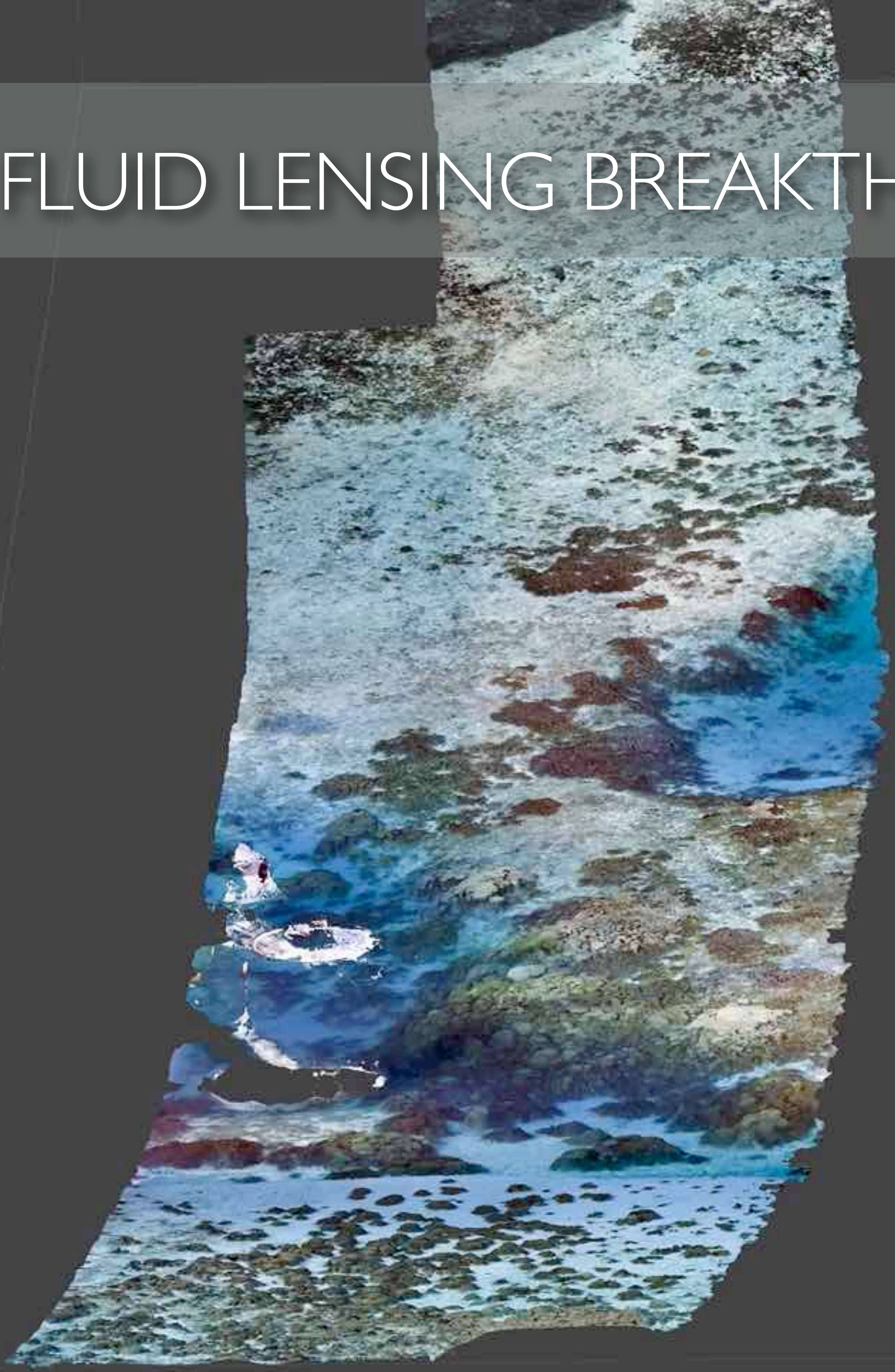
Active Learning Training of Coral Cover & Morphology Type

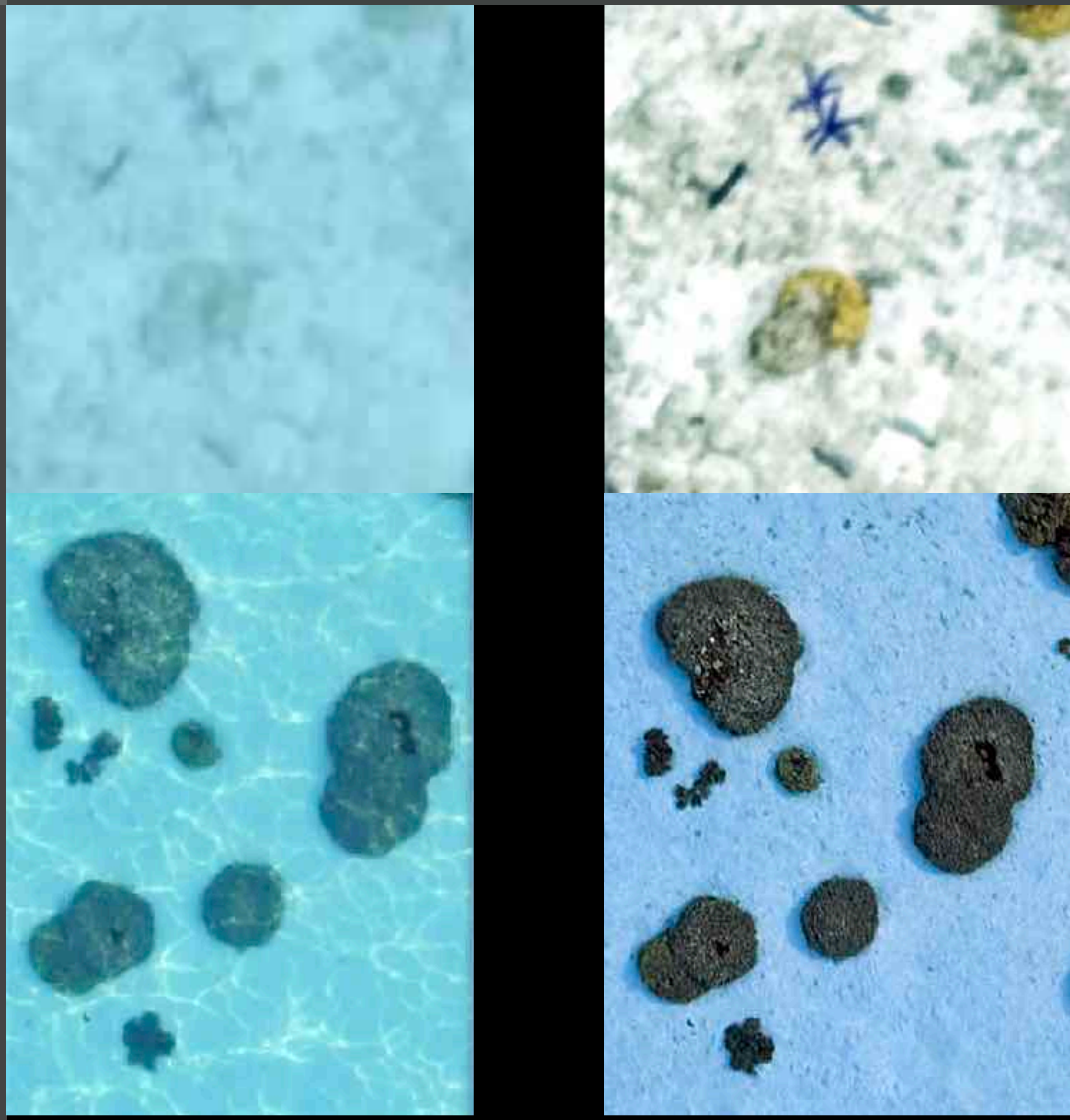
NASA IUCN NOAA NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION U.S. DEPARTMENT OF COMMERCE UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

NEW PUBLICATIONS

1) *Chirayath, V. 2020. "System and method for imaging underwater environments using fluid lensing." United States Patent and Trade Office No. 16/393,569, 2020.*
2) *Li, Alan S., Chirayath, V., et al. 2020. "NASA NeMO-Net's Convolutional Neural Network: Mapping Marine Habitats with Spectrally Heterogeneous Remote Sensing Imagery." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13 (2020): 5115-5133.*
3) *Chirayath, V. et al. 2020. "NASA NeMO-Net – A Neural Multimodal Observation & Training Network for Marine Ecosystem Mapping at Diverse Spatiotemporal Scales." IEEE Geoscience and Remote Sensing Society. In press.*
4) *Asanjan, A., Das, K., Li, A., Chirayath, V., Torres- Perez, J., and Sorooshian, S. 2020. "Learning Instrument Invariant Characteristics for Generating High-resolution Global Coral Reef Maps. 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).*
   2020. ACM, New York, NY, USA, 8 pages.
5) *Chirayath, Ved, and Alan Li. "Next-Generation Optical Sensing Technologies for Exploring Ocean Worlds-NASA FluidCam, MiDAR, and NeMO-Net." Frontiers in Marine Science 6 (2019): 521.*
6) *"Drone takes to the skies to image offshore reefs". Nature (2019) Vol 570. 545.*
7) *Current pre-release code (release_v0.1) implementation available to public (working through official NASA release): https://github.com/NASA-NeMO-Net/NeMO-Net.git*

NEW FLUID LENSING BREAKTHROUGH - 45FT DEPTH 3D IMAGE!





THANK YOU!

Supporting Shellfish Aquaculture in the Chesapeake Bay using AI for Water Quality

Stephanie Schollaert Uz (PI, NASA GSFC)
Troy Ames (Co-I, NASA GSFC)

AIST-18-007 Annual Technical Review
February 5, 2021

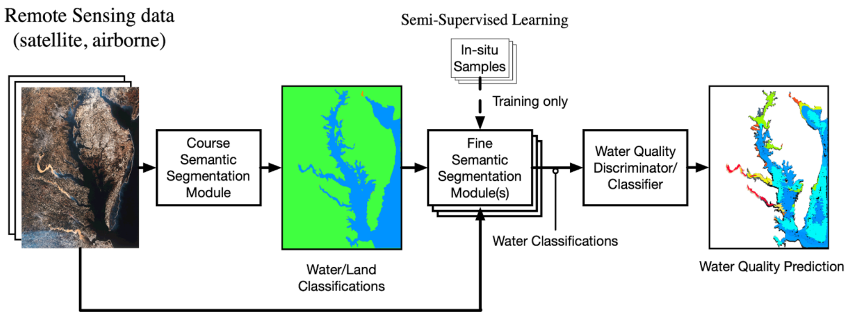
Team: Nargess Memarsadeghi, NASA GSFC
John 'Rusty' McKay, MDE
Shannon McDonnell, UMD/ESSIC and Neil V. Blough, UMD
Amita Mehta, NASA GSFC/UMBC

Supporting shellfish aquaculture in the Chesapeake Bay using AI

PI: Stephanie Schollaert Uz, NASA GSFC

Objective

- Develop a low-cost proof-of-concept artificial intelligence neural network model to detect poor water quality in the well-sampled Chesapeake Bay
- Initial performance goals are >90% accuracy for detection of poor water quality (exceeding MDE indicators, e.g. low oxygen, turbidity, fecal coliform threshold, harmful algal bloom)
- Technology includes data collection (publicly available in situ data, satellite data, hand-held optical spectroscopy, hull-mounted flow through sampling) and artificial intelligence software
- Improved capability prepares to exploit hyperspectral sensing by future PACE, GLIMR, SBG missions



Hierarchical Semantic Segmentation: a modular framework for detecting water quality from remote sensing and in situ data

Approach

Apply NASA data, science, and technology to support interagency partners (e.g. MDE, NOAA) in their operations toward the development of a decision support tool for shellfish aquaculture:

1. Collect and analyze all available in situ and remotely sensed data relevant to Chesapeake Bay water quality
2. Collect and analyze absorption and fluorescence properties of in-water constituents at hyperspectral resolution for select sites
3. Train an ACF AI to identify poor water quality that resulted in shellfish bed closure
4. Refine and validate the AI against current conditions

Co-Is/Partners: Troy Ames, GSFC; Nargess Memarsadeghi, GSFC; John McKay, Maryland Department of Environment

Key Milestones

- In situ, above-water, satellite data collection 01/20
- Data analysis and processing 01/20
- Artificial Intelligence training, tuning 12/20
- Artificial Intelligence validation (TRL 3) 06/21

$TRL_{in} = 1$ $TRL_{current} = 2$

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background

- This project helps meet R&A and Applications goals for cross-cutting science areas, namely the development of new observations beyond traditional ocean color (i.e. chlorophyll-a) from space-based assets (Ocean Biology & Biogeochemistry); promoting the use of Earth observations to monitor systems and develop management strategies (EcoForecasting; Water/Food Security; Health & Air Quality)

- Reliable information on water quality is not currently available at the space and time scales that are required for aquaculture and other resource management needs.

- Technical approach: develop an analytic center framework (ACF) to harmonize disparate in situ and remotely sensed datasets; use Artificial Intelligence (AI) to detect patterns of poor water quality not previously possible through traditional techniques

Objectives

This project is addressing several questions:

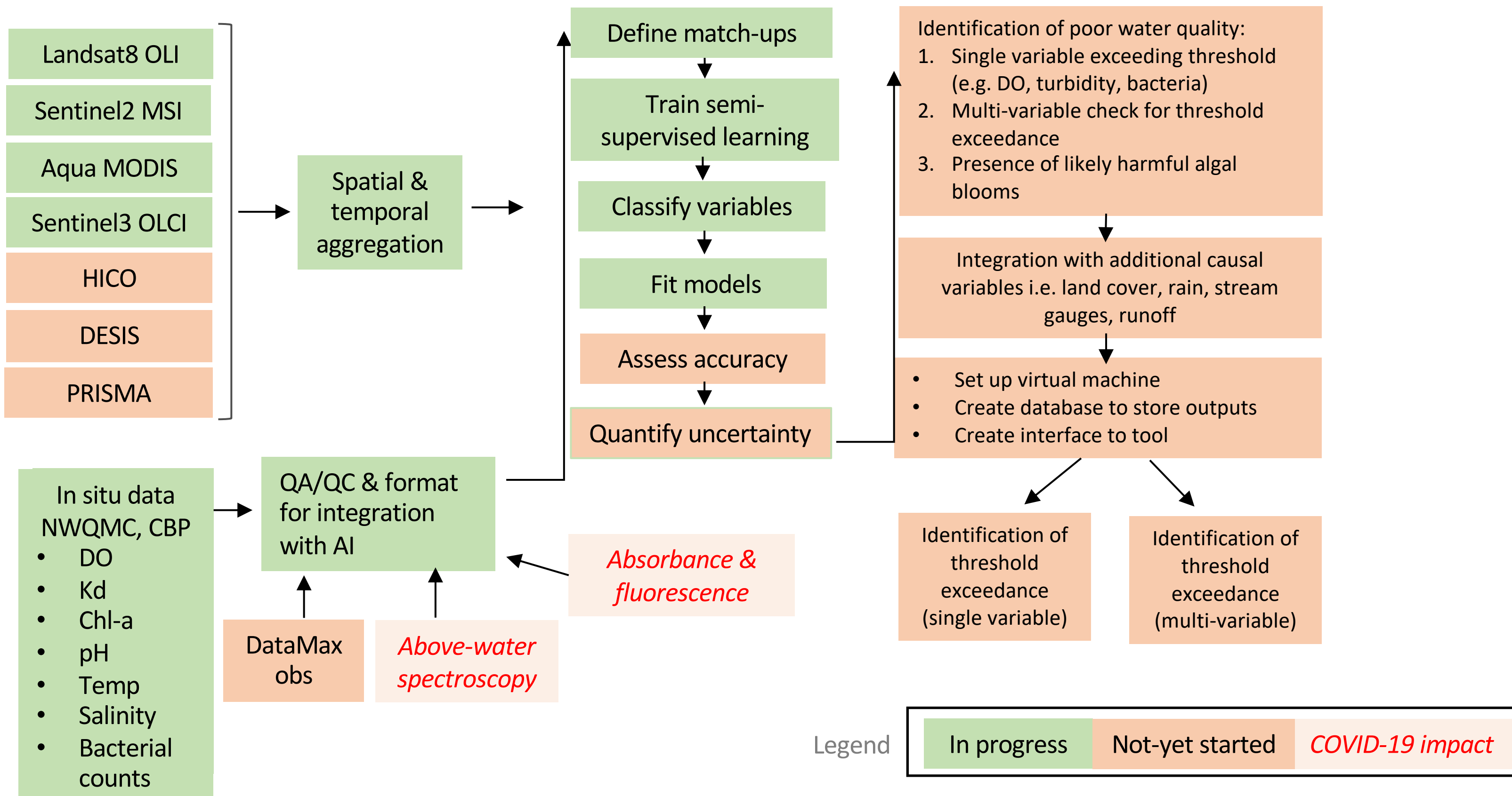
- Existing technologies and data collected in the Chesapeake Bay over several decades are being integrated and tested to determine which indicators of poor water quality we can identify via remote sensing using big data analytics and machine learning. Which indicators can we detect? What do we need to be able to detect additional indicators?

- A second component of this project is to integrate higher spectral information of inherent optical properties in cases of good and poor water quality to determine the minimum set of requirements to deliver effective remote sensing of water quality for aquaculture (i.e. critical wavelengths, spectral width, spatial and temporal resolution)?

- Many of the water quality features that impact aquaculture happen close to land at fine spatial scales where moderate-resolution satellite data lack spatial or spectral resolution. We plan to add commercial data into the system to augment the nearshore environment.

Objectives: AIST18-007 project overview

As of 1 Jan 2021



Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Technical and Science Advancements

Technical and science advancement since project kick-off: data acquired, analyzed, data segmentation begun, application formulated

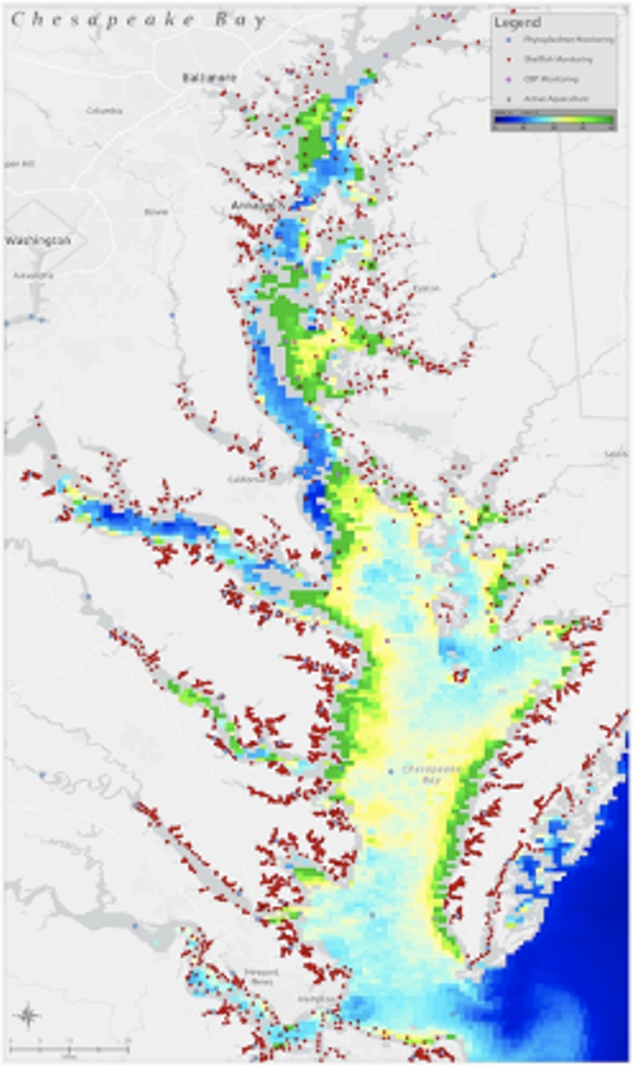
Current TRL Rational: Experiments ongoing with satellite data and synthetic data analog

Technology Readiness Level - (TRL)	Definition	Hardware Description	Software Description	Exit Criteria
1	Basic principles observed and reported	Scientific knowledge generated underpinning hardware technology concepts/applications.	Scientific knowledge generated underpinning basic properties of software architecture and mathematical formulation.	Peer reviewed publication of research underlying the proposed concept or application
2	Technology concept or application formulated	Invention begins, practical application is identified but is speculative, no experimental proof or detailed analysis is available to support the conjecture.	Practical application identified but speculative, no experimental proof or detailed analysis available to support the conjecture. Basic properties of algorithms & concepts defined. Basic principles coded. Experiments with synthetic data.	Documented description of the application or concept that addresses feasibility and benefit
3	Analytical or experimental critical function or characteristic proof-of-concept	Analytical studies place technology in appropriate context and laboratory demonstrations, modeling and simulation validate analytical prediction.	Development of limited functionality to validate critical properties and predictions using non-integrated software components	Documented analytical or experimental results validating predictions of key parameters

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Summary of Accomplishments



MODIS chlorophyll-a map from July 2, 2019 with routine sampling sites by Maryland and Virginia superimposed

- NCCS ADAPT partition created for project to exploit GPU processing and large storage
- In situ data from Chesapeake Bay Program (CBP) & National Water Quality Monitoring Council (NWQMC) acquired for Chl-a, DO, Kd, pH, salinity, Secchi depth, nutrients, water temp, plus bacteriological fecal coliform from MDE
- Multiple ocean color satellite data sets acquired

Summary of Accomplishments

Remote Sensors	Spectral Bands	Spatial Resolution	Revisit Time (Days)	Period
Aqua MODIS	36 (1 red, 1 NIR)	250m & 500m& 1km	1-2	2002-present
Sentinel 3A&3B OLCI	>15 (5 on red edge)	300m	2	2016-present
Landsat 8 OLI	4 (1 red,1 NIR)	15m & 30m	16	2013-present
Sentinel 2A&2B MSI	5 (1 red, 2 NIR, 1 red edge)	10m & 20m	5	2015-present

Current optical remote sensing of water quality parameters, e.g. chlorophyll-a, water clarity

- Level 1 satellite data processed with SeaDAS to derive Rrs, Chl-a, later Rhos to circumvent inconsistencies in atmospheric correction i.e. Rayleigh-corrected top-of-atmosphere reflectances per Wolny et al., 2020
- Later plan to explore hyperspectral HICO, DESIS, PRISMA satellite data in combination with field and lab work to prepare to exploit expanded capability of PACE, GLIMR, SBG

Summary of Accomplishments

Parameter name	Water Quality Threshold
Fecal coliform	<14 MPN median per100ml
Bacteriological *Escherichia coli*	< 410 count per 100ml
Dissolved oxygen	> 5 mg/l
Temperature	< 90°F/32°C
pH	6.5 - 8.5
Turbidity	<150 nephelometer turbidity units
Color	< 75 platinum cobalt units
Water clarity	> 13% (tidal fresh)

Water quality criteria for shellfish harvesting
http://www.dsd.state.md.us/comar/comarhtml/26/26.08.02.03-3.htm

- Limited sampling prior to COVID-19 and no poor cases, thus targeting hypoxia first as a large scale, seasonal problem related to physical drivers

Summary of Accomplishments

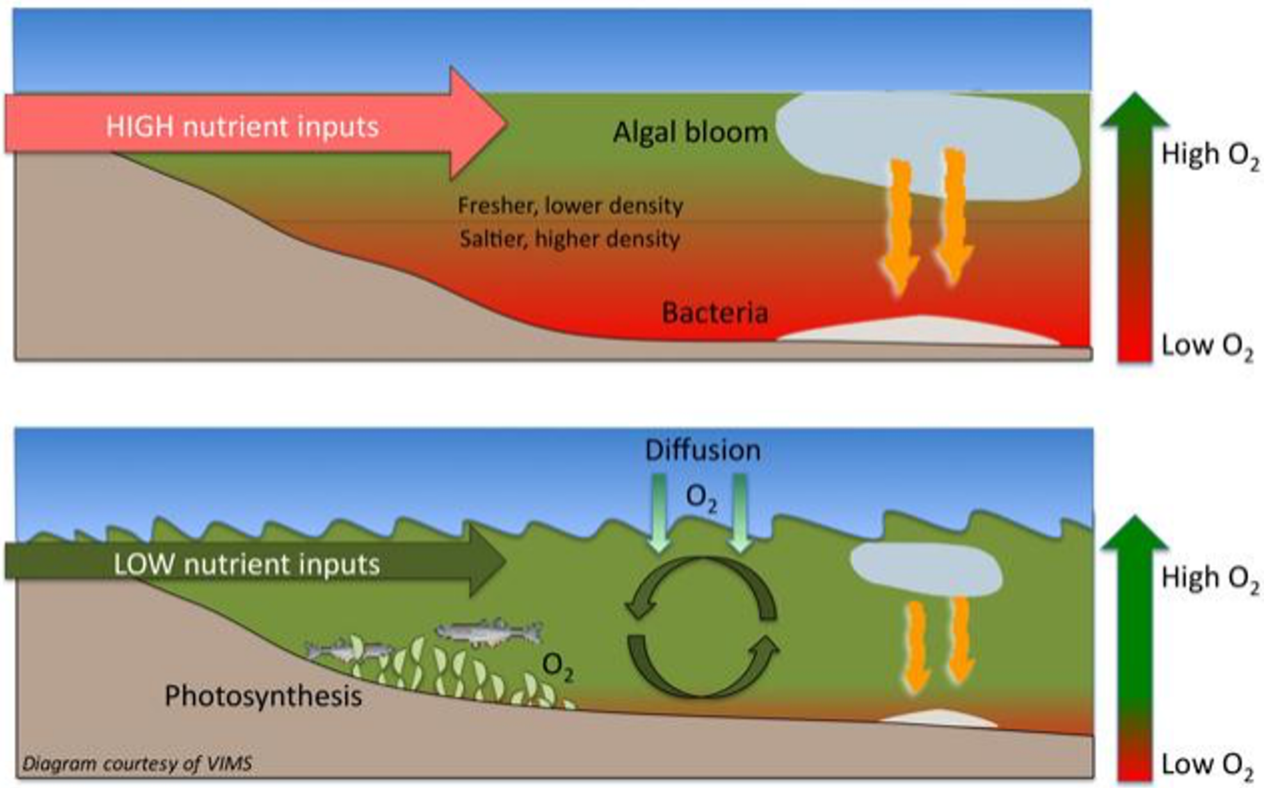
Parameter name	Water Quality Threshold
Fecal coliform	<14 MPN median per100ml
Bacteriological *Escherichia coli*	< 410 count per 100ml
Dissolved oxygen	> 5 mg/l
Temperature	< 90°F/32°C
pH	6.5 - 8.5
Turbidity	<150 nephelometer turbidity units
Color	< 75 platinum cobalt units
Water clarity	> 13% (tidal fresh)

Water quality criteria for shellfish harvesting
http://www.dsd.state.md.us/comar/comarhtml/26/26.08.02.03-3.htm

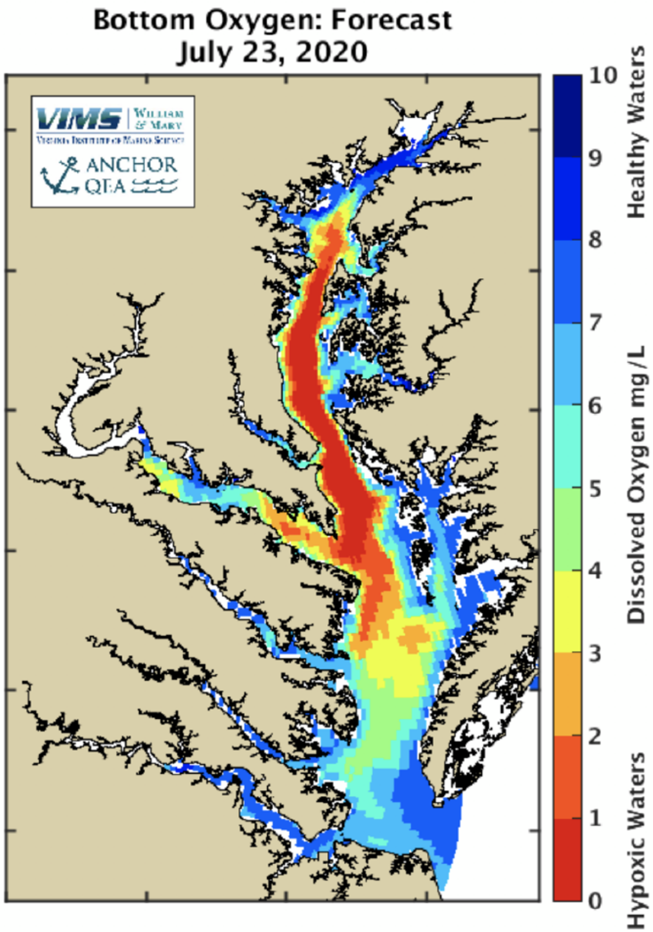
- Limited sampling prior to COVID-19 and no poor cases, thus targeting hypoxia first as a large scale, seasonal problem related to physical drivers
- Virginia Institute of Marine Science (VIMS) shared output from geophysically forced hypoxia model at 600m, 20 level, daily resolution. Model since 1984, we're using 2002-present (ocean color satellite era)
- Using hypoxia model output as synthetic data analog for training, can tune it with in situ data

Summary of Accomplishments

- Establish proof-of-concept for using optical remote sensing in ML for water quality
- Low Dissolved Oxygen (Hypoxia) is large scale, seasonal problem related to biological and physical mechanisms, i.e. phytoplankton blooms, bacteria, stratification
- Although hypoxia is not visible in remote sensing, training AI to recognize patterns that lead to it will demonstrate its feasibility to identify water quality
- Targeting its identification first by training AI with Rhos and other precursors in combination with VIMS model output trained with CBP O_2 profiles



Summary of Accomplishments



Blue -> high bottom oxygen

Yellow/green -> marginal oxygen

Red -> very low bottom oxygen (hypoxia)

Feature variables:

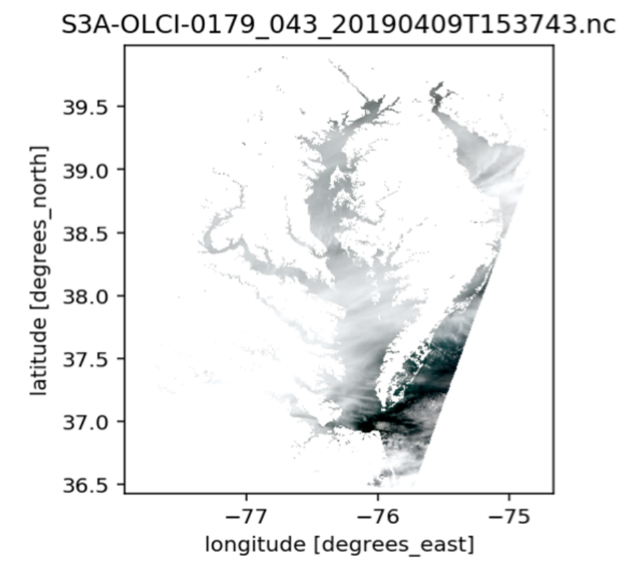
- **Satellite**: Rhos to indicate organic matter
- **ERA5**: wind (u & v)
- **VIMS**: currents (u, v, w), water temperature (T), salinity(S)
- **Ancillary**: day of year for seasonal variations

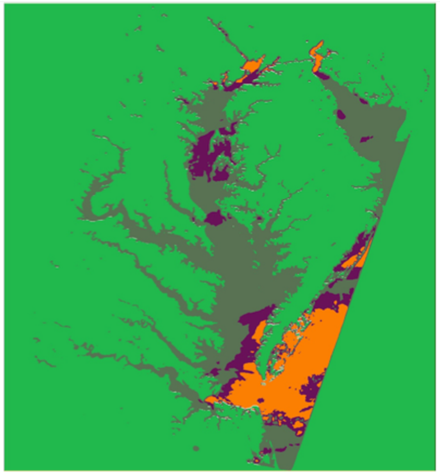
Label variable: DO (from VIMS and CBP in situ data)

- 3-D array prediction for a region

Summary of Accomplishments

- Developed pipeline to pull updated wind data from the National Data Buoy Center
- Completed data pipeline for preprocessing and merging in situ water quality data
- Developed tools to collate remote sensing and in situ data for time, location, and quality
- Completed design and implementation of a ML framework for S3 OCLI spectra data to assess water quality parameters for a specific pixel
- Completed ML design for a feature segmentation framework for S3 OCLI (example on right)
- Defined requirements for using ML for mapping hypoxia over time using wind, tides, stream flow, and bathymetry

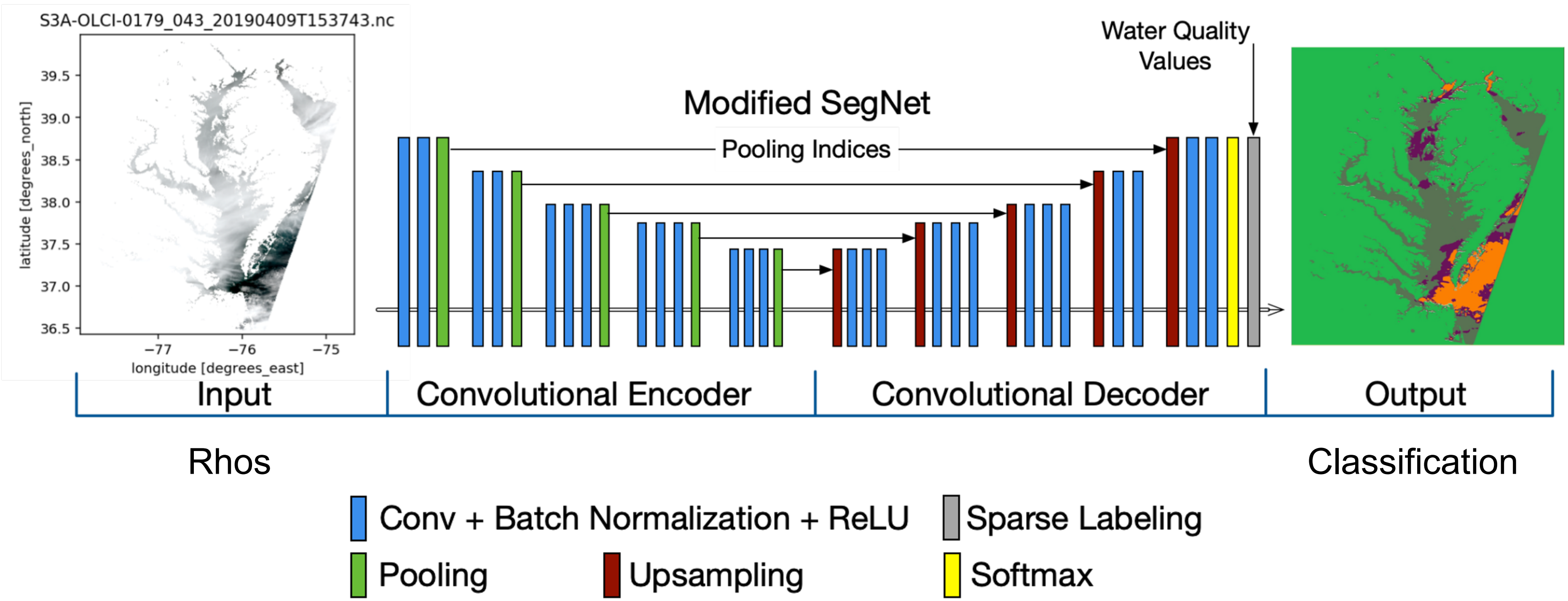




Segmentation of Rhos bands

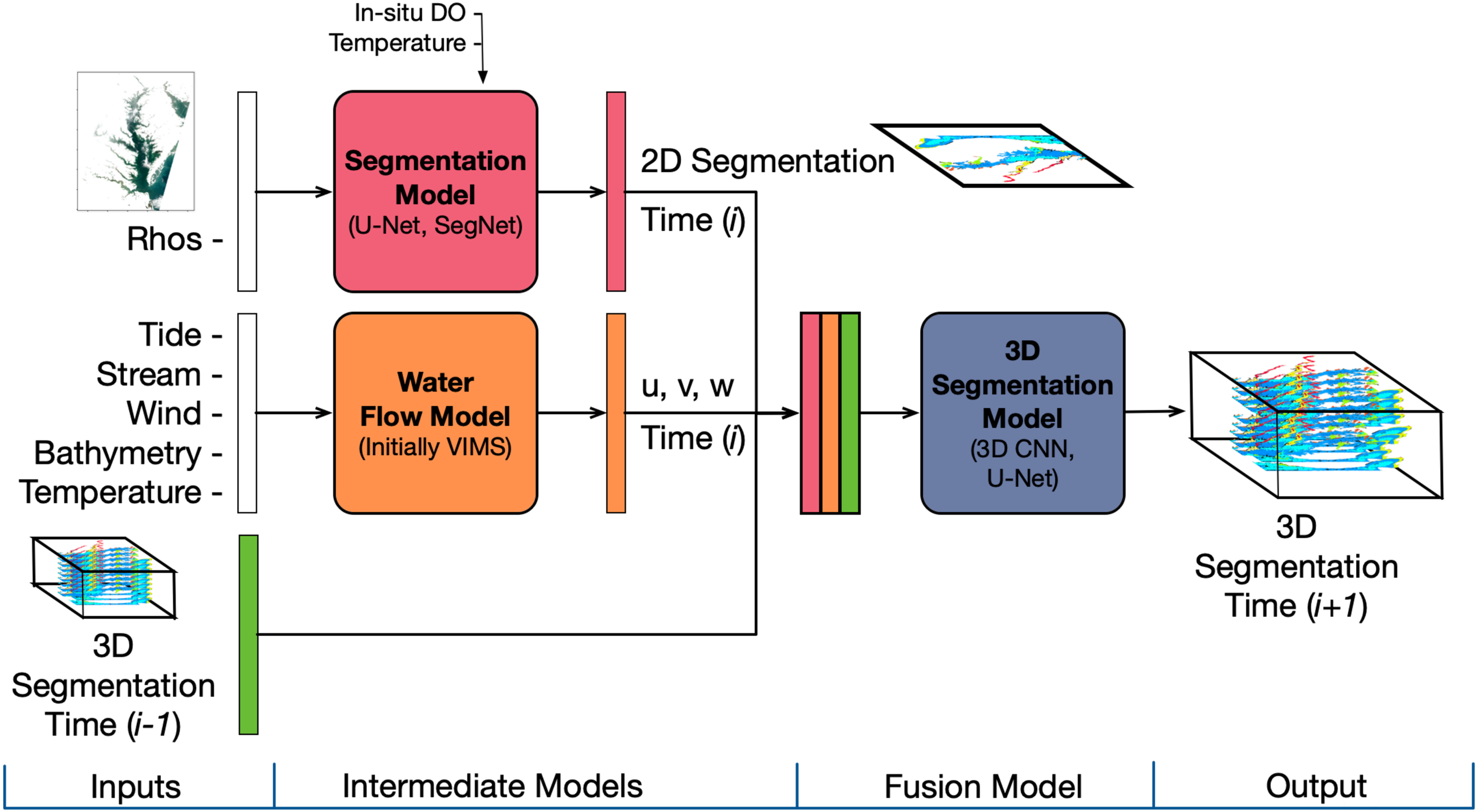
Summary of Accomplishments

Water Quality Parameter Segmentation (unsupervised)
Segmentation Model



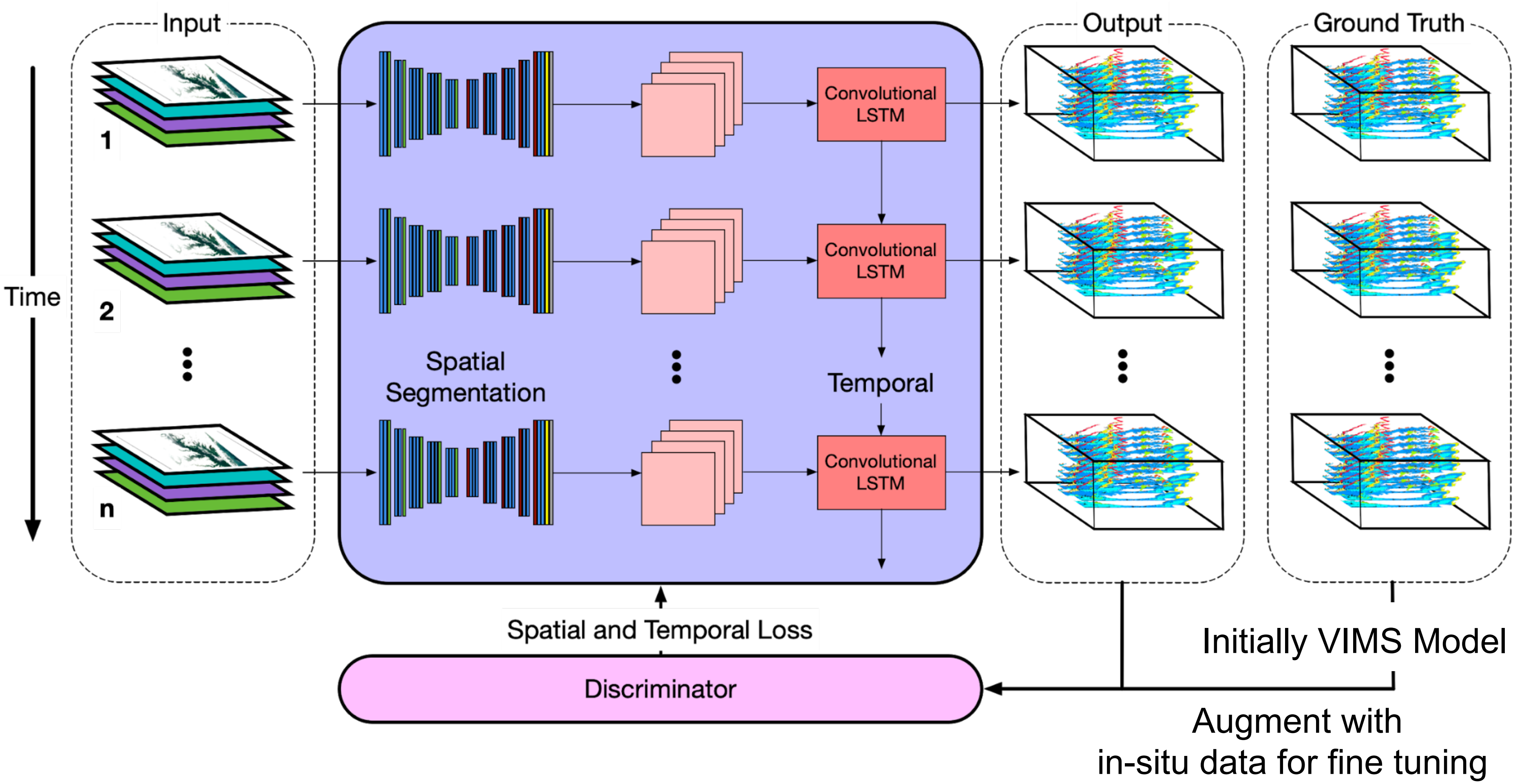
Summary of Accomplishments

Composite Hypoxia ML Model



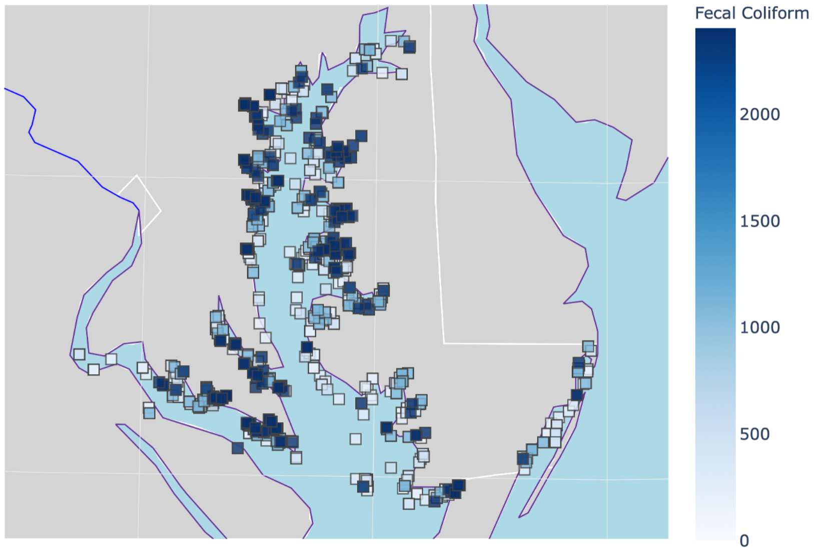
Summary of Accomplishments

Spatial/Temporal Segmentation Model



Future Plans

- Complete Semantic Segmentation pipeline modules
  - Feature matching utilizing weakly annotated data (i.e., in situ measurements)
- Implement additional modules for the hypoxia prediction ML framework
- Develop methods for ingesting and preparing VIMS data for training
- Create training, validation, and test data sets for hypoxia



- Evaluate applicable data augmentation strategies for training
  - GAN for generating extra training examples
- Resume data collection and analysis: target poor water quality; distinguish pertinent amino acids, e.g. tyrosine/tryptophan, brighteners. Analyze spectral ratios of phytoplankton and bacteria regions within complex CDOM spectra
- Develop identification of additional phytoplankton pigments, likely harmful algal bloom (HAB)

Summary of Accomplishments and Future Plans

- Historical and current in situ and satellite data collected & hypoxia model output for ML training
- AI proof-of-concept objective defined. Training and testing ongoing
- Currently running parallel testing for different satellite platforms
- Data fusion will be explored for different satellites and platforms. Quality label to score matchups (<1hr best; 1-3hr good; 3-9hr fair; 9-24hr poor)
- DataMax collection by MDE will resume in spring and provide continuous underway measurements to detect phytoplankton pigments, high nutrient concentration, contamination, effluent, fresh or brine discharges, spills. We plan to target poor water quality cases
- NOAA Hand-held hyperspectral sensor (hypergun) will provide point source spectral information to augment satellite data with above-water optical properties
- Once lab work resumes, absorbance and fluorescence measurements will be analyzed for different cases, poor and good, in order to train AI objective for exploiting higher spectral characteristics
- In situ data will be used to tune, test, and validate ML, assess accuracy, uncertainties

Presentation Contents

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

Publications

Journal / Conference Papers

IGARSS 2020 manuscript presented as an oral presentation
#3587, 'SUPPORTING AQUACULTURE IN THE CHESAPEAKE BAY USING ARTIFICIAL INTELLIGENCE TO DETECT POOR WATER QUALITY WITH REMOTE SENSING'

Wolny, J.L., M.C. Tomlinson, S. Schollaert Uz, T.A. Egerton, J.R. McKay, A. Meredith, K.S. Reece, G.P. Scott, and. R.P. Stumpf, 2020, Current and Future Remote Sensing of Harmful Algal Blooms in the Chesapeake Bay to Support the Shellfish Industry, *Front. Mar. Sci.,* doi:10.3389/fmars.2020.00337

Dissertations

McDonnell, S.M. *in prep.*

Other

List of Acronyms

CBP	Chesapeake Bay Program
CDOM	Colored Dissolved Organic Matter
CNN	Convolutional Neural Network
DESIS	DLR (German Space Agency) Earth Sensing Imaging Spectrometer
GLIMR	Geostationary Littoral Imaging and Monitoring Radiometer
HAB	Harmful Algal Bloom
HICO	Hyperspectral Imager for the Coastal Ocean
LSTM	Long Short Term Memory
MODIS	Moderate-resolution Imaging Spectrometer
MSI	Multispectral Imager
NWQMC	National Water Quality Monitoring Council
OLCI	Ocean and Land Color Instrument
OLI	Operational Land Imager
PACE	Plankton Cloud ocean Ecosystem
PRISMA	(Italian) Hyperspectral Precursor of the Application Mission
Rhos	Top-of-atmosphere reflectance minus Rayleigh
Rrs	Remote sensing reflectance
SBG	Surface Biology and Geology
SST	Sea-Surface Temperature
S2	Sentinel-2 A&B
S3	Sentinel-3 A&B

NASA Evolutionary Programming Analytic Center (NEPAC) for Climate Data Records, Science Products and Models

John R. Moisan (PI, NASA/GSFC Code 610.W)
Carlos del Castillo (Co-I, NASA/GSFC Code 616)

AIST-18-0031 Annual Review Meeting
5 February 2021

Team listing:
Guoqing Wang (SSAI)
Yuping Liu (SSAI)
Code 606 Support:
Mark Caroll, Roger Gill, Calib Spradlin

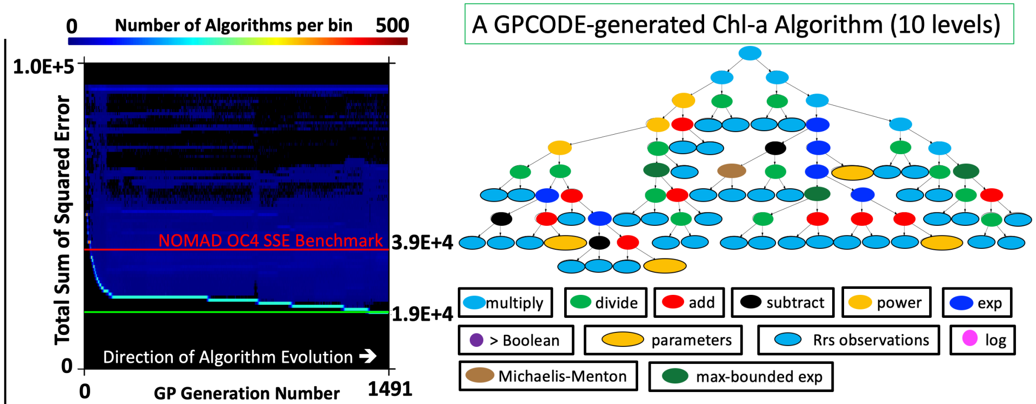
NASA Evolutionary Programming Analytic Center (NEPAC) for Climate Data Records, Science Products and Models

PI: John R. Moisan / NASA/GSFC

Objective

Bring Genetic Programming (GP) into a state of general use by NASA scientists and engineers. NEPAC will be an Analytic Center Framework (ACF) that links satellite and *in situ* data sets to a GP code with bootstrapping and performance metrics to objectively evolve empirical algorithms for satellite products. A Graphical User Interface (GUI) will escort users through a simple self-described procedure for choosing data sets, optimization and GP operational criteria, and specific computational assets (*e.g.* NASA HEC, ESTO or other Cloud).

Performance goals are to: (1) create chlorophyll a (Chl-a) algorithms with a >50% improvement over present algorithms, (2) eliminate present over/underfitting of low/high Chl-a estimates, (3) quantify algorithm uncertainties and errors, (4) yield an improved Chl-a time series across multiple satellites sensors, and (5) prototype NEPAC to the larger NASA community of algorithm developers.



GP created algorithms outperform the classic OC4 Chl-a algorithm. A time sequence of SSE histograms from the algorithm population (upper left) shows the rapid evolution of Chl-a algorithms (*e.g.* upper right) that outperform the OC4 algorithm. NEPAC's long term goal is to bring this capacity into general NASA use.

Approach

Design the NEPAC's ACF system and interface based upon target community user requirements.

Develop a web-enabled GUI for NEPAC front end.

Develop APIs to link data sets to data selection and uploads, and for submitting GP runs to the HEC and collect post-run analysis.

Expand GPCODE capability by:

1. Insertion of bootstrap sampling to limit algorithm overfit and support uncertainty estimation,
2. Upgrading optimization performance metrics with methods that utilize data error and variances, and
3. Demonstrating GP's utility by evolving improved Chl-a algorithms to the targeted science community.

Co-Is/Partners: Dr. Carlos Del Castillo (co-I)

Key Milestones

- **Completed ACF flowchart, requirements, testing plan 10/20**
- Upgrade GP, create APIs and remote HEC access. 01/21*
- Complete GUI interface and testing components 10/21
- Prototype validation (TRL 6) 12/21

*Modified milestones to eliminate remote HES access in order to create a 'containerized' NEPAC application that allows any user to apply NEPAC using other HEC infrastructure.
API's are under development with Code 606 partnership.

TRL_{in} = 2 $TRL_{current}$ = 4/5

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background and Objectives

- This project is focused on improving the satellite algorithm for ocean chlorophyll-a, the key biological state variable for ocean and inland water ecosystem assessment, historically the first-order proxy for phytoplankton biomass estimates and necessary inputs for contemporary carbon-based productivity estimates, making Chl-a observations a critical element for global phytoplankton climate assessment. The key Earth Science Programs that this project supports are the Carbon Cycle (Laura Lorenzoni) and Ecosystem Forecasting (Woody Turner).

- **Objectives**: To develop an Analytic Center Framework (ACF), called the NASA Evolutionary Programming Analytic Center (NEPAC), for rapid formulation of satellite algorithms for Chl-a. NEPAC will generate Chl-a algorithms with reduced uncertainties and the capability to anneal Chl-a estimates across multiple Ocean Color (OC) satellite data sets to support coherent Chl-a Climate Data Records (CDRs).
- **Technology**: The core technology element is a Genetic Programming (GP) application that generates equations/algorithms. A Graphical User Interface (GUI) will provide user access to the GP toolbox and allow for selection of data sets and GP control inputs and connect data and GP code to high end computer resources.
- **Science Goals**: The overall science goal is to use the improved Chl-a algorithm, with lower errors, bias and uncertainty, to generate global Chl-a estimates that can be used to analyze the long-term and climate-scale changes in ocean biomass and ocean carbon cycling dynamics, and key science questions regarding the sensitivity of ocean ecology.

Presentation Contents

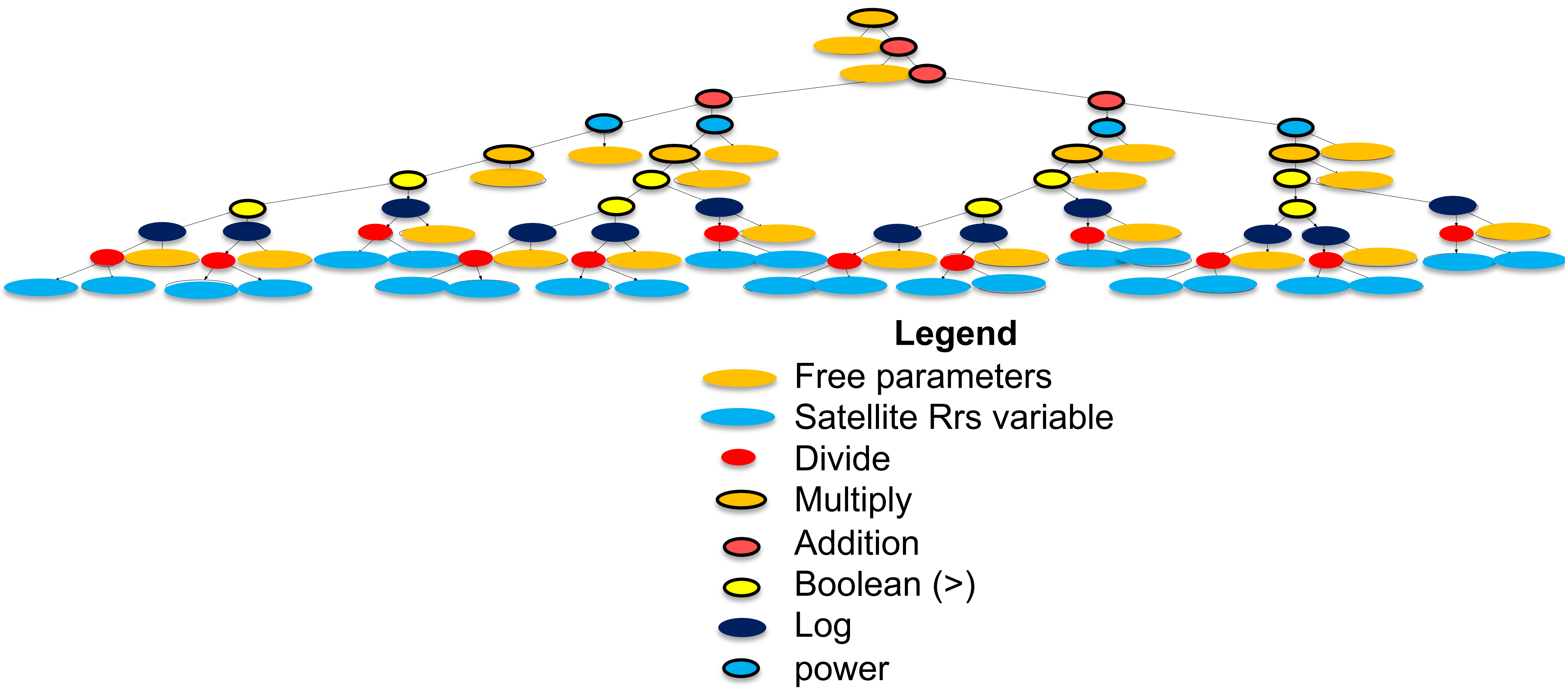
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Technical and Science Advancements

- The project has made progress in each of the AC's components.
  - The front-end user interface is presently carrying out testing of a process that takes user-supplied in situ observations of chlorophyll a in situ to mine the OC_DAAC and other data archives to generate a training data set for GPCODE runs.
  - The GPCODE has completed installation of several modifications that are aimed to provide improvements in satellite algorithms.
    - First, there is now an option of using Mean Absolute Percent Error (MAPE) as a metric for algorithm evolution. This has led to providing improvements in generating algorithms for observations that have not first been log-transformed—as the present satellite algorithms are. And the solutions obtained using the MAPE metric have not impacted to resulting SSE improvements.
    - Second, we have inserted the option of using Bootstrapping to the data sets in order to suppress potential overfitting of the algorithms and to foster the development of sparse or 'non-bloated' algorithms. This modification is still being tested.
    - Third, the trees populations are now being analyzed to track the entropy of each of the individual trees in order to test some dynamic controls for evolution that might yield better performance in the evolutionary process. The idea is to insert a control that enhances tree diversity so that there is vibrant evolution.
  - A suite of post-run analysis programs have been created. One specifically takes the resulting GPCODE trees and converts them into usable Fortran/C/Matlab code.
- The results from the test runs are showing that the modifications are yielding improved satellite algorithms. We are in the process now of interfacing with the OCSSW to complete the NEPAC tests.

Example binary tree for OC4 Chlorophyll a algorithm

Presently used chlorophyll-a satellite algorithms are 11-level trees with 21 parameters, 5 of which are tuned to a data set. [Note: the other 16 define the log (10) scales to which the observations are transformed.]

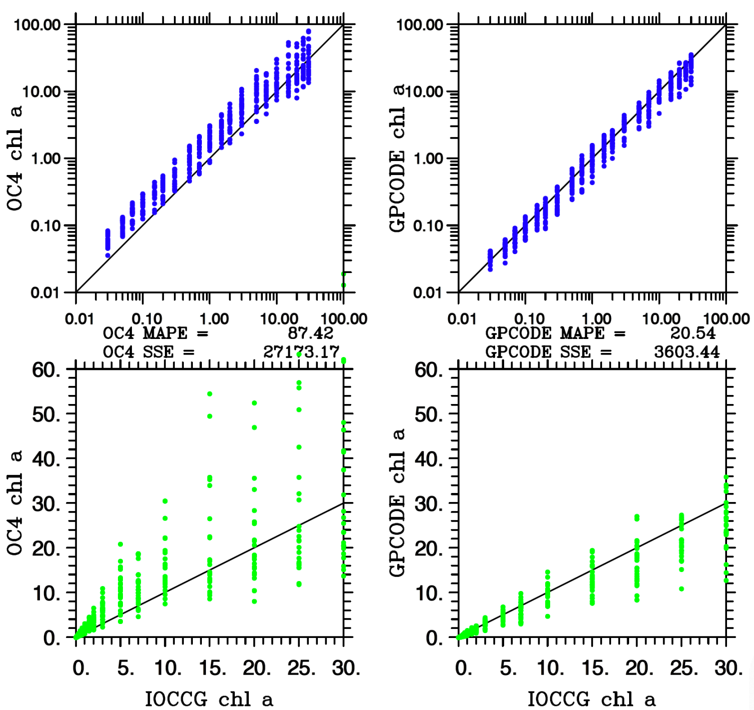


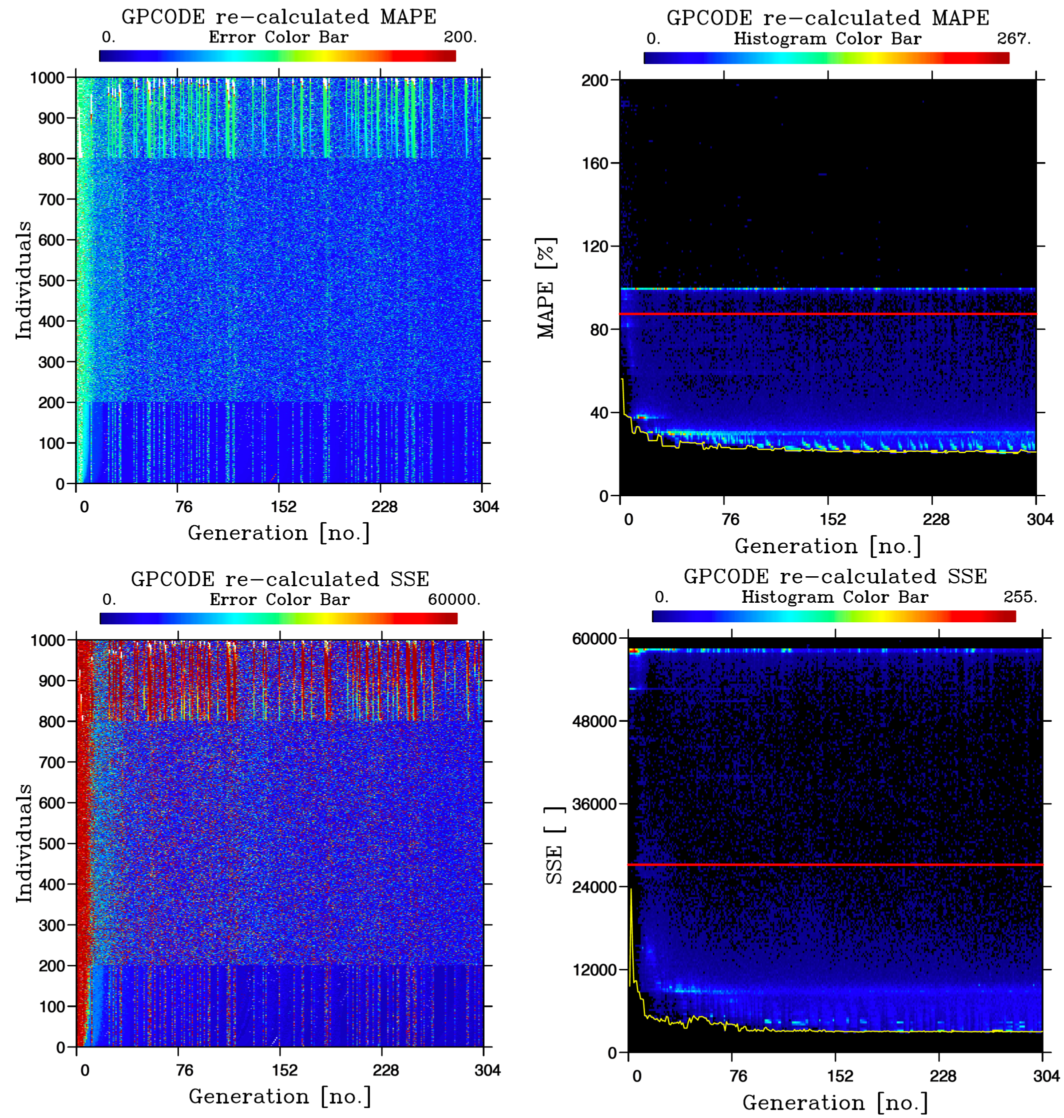
Satellite OC Algorithm
SSE: 27173.17; MAPE: 87.42%

An Example GPCODE Algorithm
SSE: 3603.44; MAPE: 20.54%

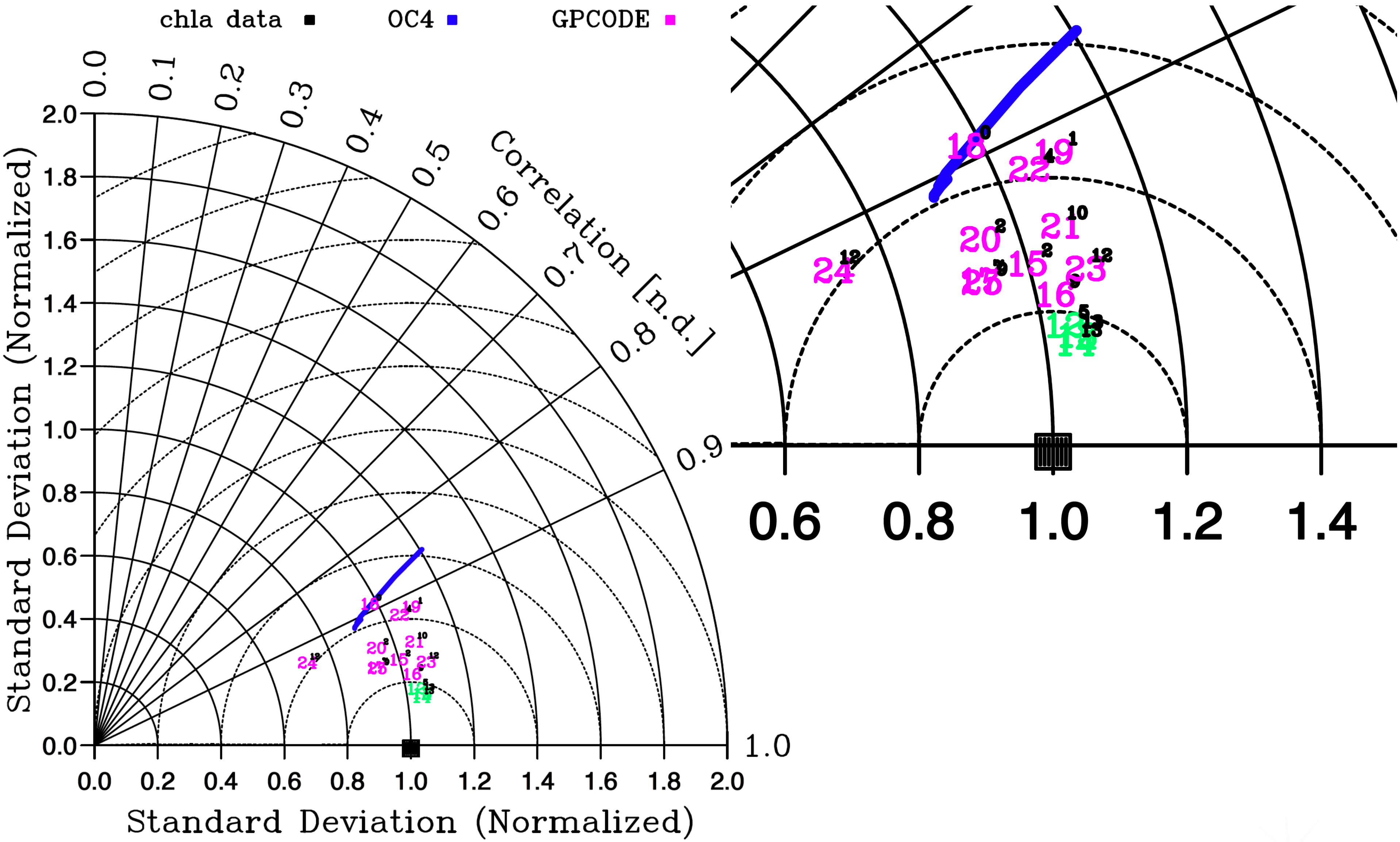
Case 12: GP_Tree_All_SSE

- n_GP_Generations=304
- n_GP_Individuals=1000
- MAPE metric
- No Bootstrapping (BS)
- # of parameters in best tree: 9
- Best SSE Tree (gen. = 238 , ind. = 1)
- Best MAPE Tree (gen. = 206 , ind. = 614)





GPCODE algorithms using IOCCG data



Presentation Contents

Summary of Accomplishments and Future Plans

- Provide a summary of the current state
- All of the NEPAC components are in various stages of development and testing.
- It has been decided that the complete NEPAC application should be delivered to the community as a 'containerized' application that can be ported onto any user's computer infrastructure. This decision came from the simple fact the NASA's IT access continues to be a moving target and having a broad use of NEPAC in the science community requires that it be user-friendly.
- Two papers (one published, one in review) and a published book chapter have been completed.
- We have been able to obtain some highly successful chlorophyll algorithms during the development tests, which is a very signal that this effort will have high return on investment.
- The front-end and development of the GUI and Containerization will make up a large part of the work in the coming year. In addition, we will be completing the steps needed to insert the tree-generated C-codes into the OCSSW application.
- Jupyter Notebook is being used to develop and test the process scripts for the NEPAC architecture to provide early beta-testing of the software applications.
- We are working on a suite of NEPAC runs that are using a variety of satellite data sets in order to develop a process for creating an algorithm that can be used for Climate Data Record time series estimates.
- Still working to incorporate the Maximum Probability Regression technique into the GPCODE.

Presentation Contents

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

ESTO
Earth Science Technology Office

Publications

- Journal / Conference Papers

Two papers that have made demonstrated the use of using chlorophyll algorithms to estimate phytoplankton functional types and their observed trends in the coastal ocean.

2020: Friedland, K. D., **J. R. Moisan**, A. Maureaud, D. Brady, A. Davies, S. Bograd, R. Watson, and Y. Rousswau, Trends in phytoplankton communities within large marine ecosystems diverge from the global ocean, *Canadian J. Fish and Aquatic Sci.*, In Review.

2020: Friedland, K. D., R. E. Morse, N. Shackell, J. C. Tam, J. L. Morano, **J. R. Moisan**, and D. C. Brady, Changing physical conditions and lower and upper trophic level responses on the US Northeast Shelf, *Frontiers in Marine Science*, https://doi.org/10.3389/fmars.2020.567445.

- Book Chapter

A review chapter presenting an overview on the techniques of estimating pigments from satellite measurements.

2020: **Wang, G.**, and **J. R. Moisan**, Remote Sensing of Phytoplankton Pigments, In "Plankton Communities", Leonel Pereira (ed.), Open Access, IntechOpen, London, ISBN 978-1-83968-609-2.

CAPRi:
Cloud-based Analytic Framework for Precipitation Research

AIST-18-0051
Group Annual Technical Review
Award Number: 80NSSC20K0205

February 5, 2021

PI John Beck -The University of Alabama in Huntsville (UAH)
Co-I Todd Berendes, Anita LeRoy, Geoffrey Stano – UAH
Patrick Gatlin – Marshall Space Flight Center, NASA

ESTO
Earth Science Technology Office

itsc
INFORMATION TECHNOLOGY AND SYSTEMS CENTER

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

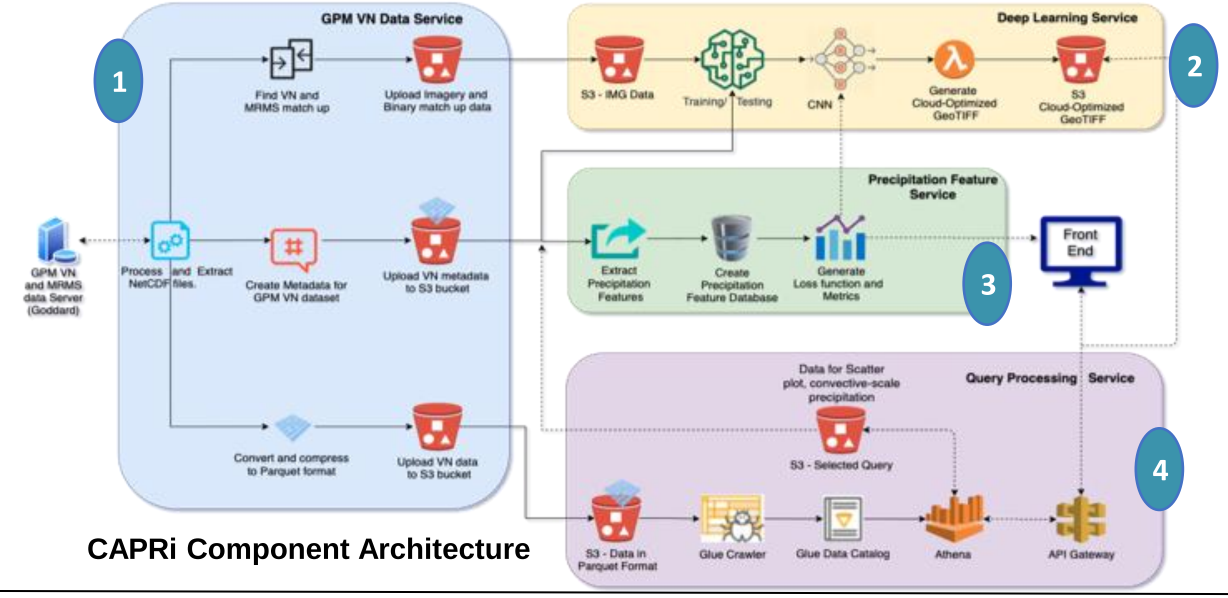
CAPRi: Cloud-based Analytic Framework for Precipitation Research

PI: John Beck, The University of Alabama in Huntsville

Objective

Develop a cloud-based integrated platform for conducting Deep Learning research to open new frontiers for extracting knowledge from the vast amounts of data generated by NASA's Earth observation systems.

- Adapt and extend cloud-based serverless technologies prototyped in the AIST-2016 VISAGE project to implement a Cloud-based Analytical Framework for Precipitation Research (CAPRi).
- Provide tools and services for accessing, analyzing, and visualizing GPM VN data. *(Figure, Components 1, 4)*
- Integrate Deep Learning models into CAPRi for super-resolution of GPM DPR radar data. CAPRi services will be developed to select and prepare training and test data. *(Component 2)*
- Use CAPRi services and the results from the super-resolution CNN to create a 3D precipitation features demonstration database. *(Component 3)*



CAPRi Component Architecture

Approach

CAPRi will support Deep Learning from space-based precipitation observations. Serverless computing tools from Amazon Web Services (AWS), including step and lambda functions, S3 object storage, and the Athena stateless query service will be used in conjunction with a Deep Learning Network. This approach provides flexible scaling to support on-demand data discovery, rendering, and analytics at minimal cost. One challenge will be the sheer volume and file size of the data for serverless technologies to handle.

Co-Is/Partners: Patrick Gatlin, NASA/MSFC; Todd Berendes, Anita LeRoy, Geoffrey Stano UAH

Collaborators: W. Petersen, NASA-MSFC

Key Milestones

Ingest GPM-VN data in CAPRi	**Completed**
Develop Deep Learning training data	**Completed**
Develop Real-time Data Querying Service	**Completed**
Develop Storage Rules and Procedures	**Completed**
Develop Metrics for Super-Resolution Results	02/21
Update CNN to process 3D Data	04/21
Create 3D Convective Scale Precip. Features	09/21
Test and evaluate CAPRi functionality	11/21
Final review and report	02/22

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

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background

We are developing a **Cloud-based Analytic Framework for Precipitation Research** (**CAPRi**) to provide users with tools for on-demand Global Precipitation Mission Validation Network (GPM VN) data querying, fusion, sub-setting, extraction, and analysis integrated with Deep Learning architectures. A part of this solution will include the ability to generate GPM training data for researchers to use within Deep Learning models.

- Base technology will leverage previous work by Co-I Berendes and Co-I Gatlin with the AIST-2016 **VISAGE** project.
- CAPRi will extend earlier VISAGE work to support **Deep Learning** in a cloud-based serverless platform.
- CAPRi technology will expand on PI Beck's research in **super-resolution** of remotely sensed images and adapt lessons learned from a successful prototype using GPM Dual-frequency Precipitation Radar (DPR) gridded data as the target for developing a higher resolution product
- A science use case will access, process, and analyze GPM VN datasets from CAPRi for **identifying and extracting three-dimensional 3D convective-scale precipitation** based on previous work by Co-I LeRoy.

Objectives

Develop a cloud-based integrated platform for conducting Deep Learning research to open new frontiers for extracting knowledge from the vast amounts of data generated by NASA's Earth observation systems.

- **Adapt and extend** cloud-based serverless technologies prototyped in VISAGE to implement a Cloud-based Analytical Framework for Precipitation Research (CAPRi).
- Integrate Deep Learning software and models into CAPRi for **super-resolution of GPM DPR radar data**.
- Develop CAPRi services and tools that will allow researchers to select and prepare **training and test data** for Deep Learning.
- Use CAPRi services and the results from the super-resolution CNN to create a **3D precipitation features** demonstration database.

Key technical challenges:

Large volume of training data for the CNN development
Temporal alignment of data with diverse time scales and resolutions
Management and cost of cloud resources

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

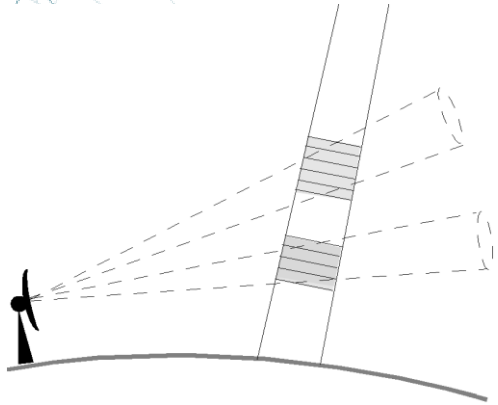
Current State – Cloud Technologies

- **Cloud Technologies** are being used in many scientific applications including the use of precipitation data.
- Advanced cloud databases, lambda functions, storage, and visualization tools are readily available. However, they can be difficult to use by most researchers.

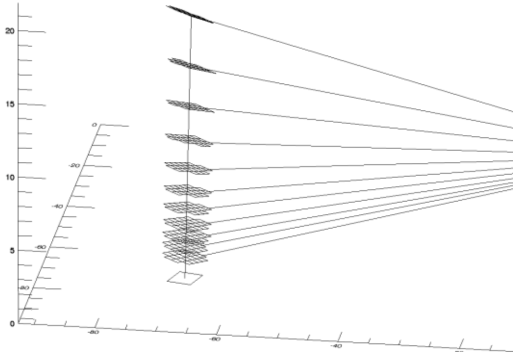
CAPRi is expanding the state of knowledge in Cloud Technologies for precipitation data by*: *(1) providing new solutions for real-time querying of large data sets in a serverless environment, (2) developing new methods for generating Deep Learning Training data on the fly, and (3) by providing an easy-to-use user interface for analysis and visualization of the data. Results will have wide application to other types of NASA datasets has the potential to impact the state of knowledge in other disciplines that require large diverse training data for Deep Learning. To demonstrate this technology, we propose to use CAPRi services for identifying **convective scale 3D precipitation features**.

Current State – GPM Validation Network (VN) Data

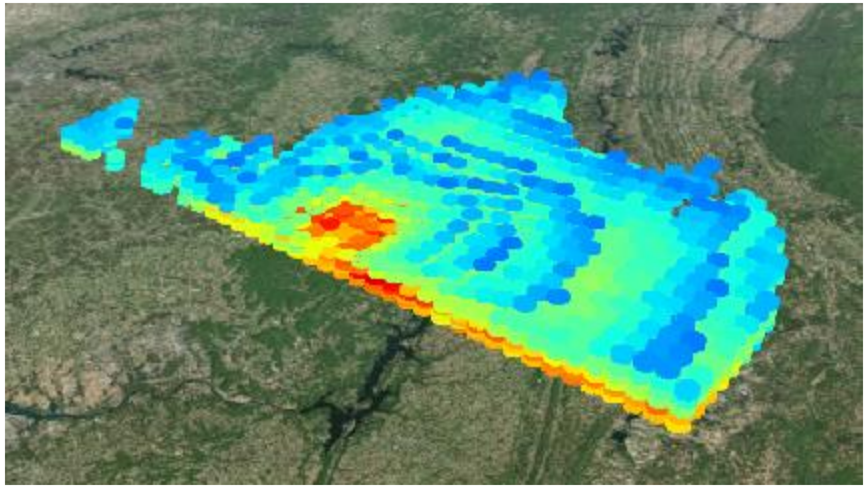
- VN matchup is a GPM footprint-based geometric match between ground radar (GR) acquired at nexrad (and other) sites and the GPM satellite overpass date/time
- GPM gates are averaged within intersecting GR elevation sweeps
- GR gates are averaged within GPM ray footprints at intersections
- 3D cylindrical volumes represent intersection of GPM rays and GR scans within the GPM vertical column (footprint)



DPR gate averaging at GR sweep intersections. The values of the individual DPR gates are averaged over the vertical extent of the GR sweeps, resulting in two matching volumes for the single DPR ray shown in this case.



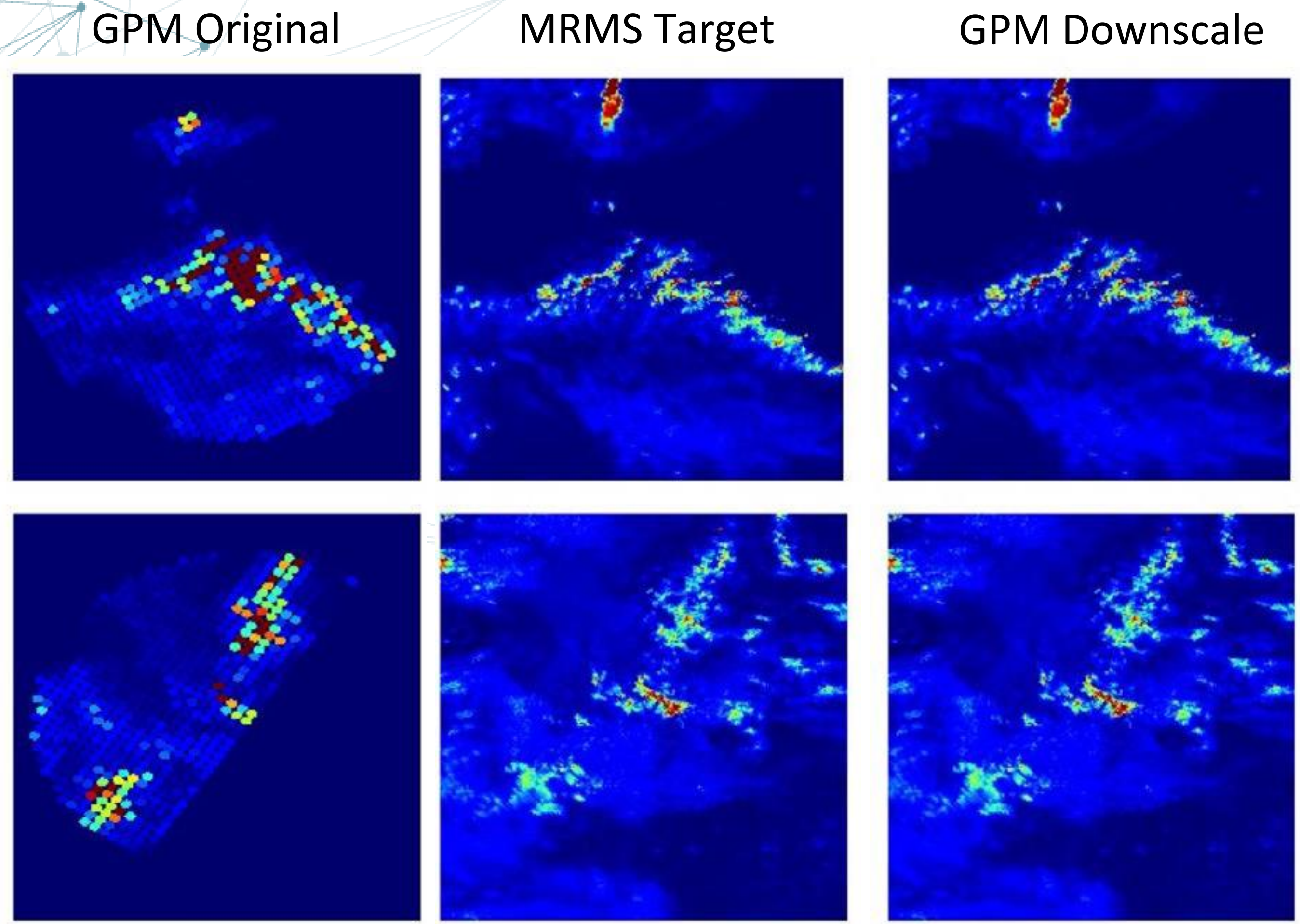
GR volume matching to DPR. The "waffle" areas show the horizontal outline of GR gates mapped to the DPR ray for each individual elevation sweep of the ground radar.



3D display of VN matchup volumes of DPR reflectivity

Current State – Deep Learning and Data Fusion

GPM VN super-resolution of surface rain rate data using MRMS



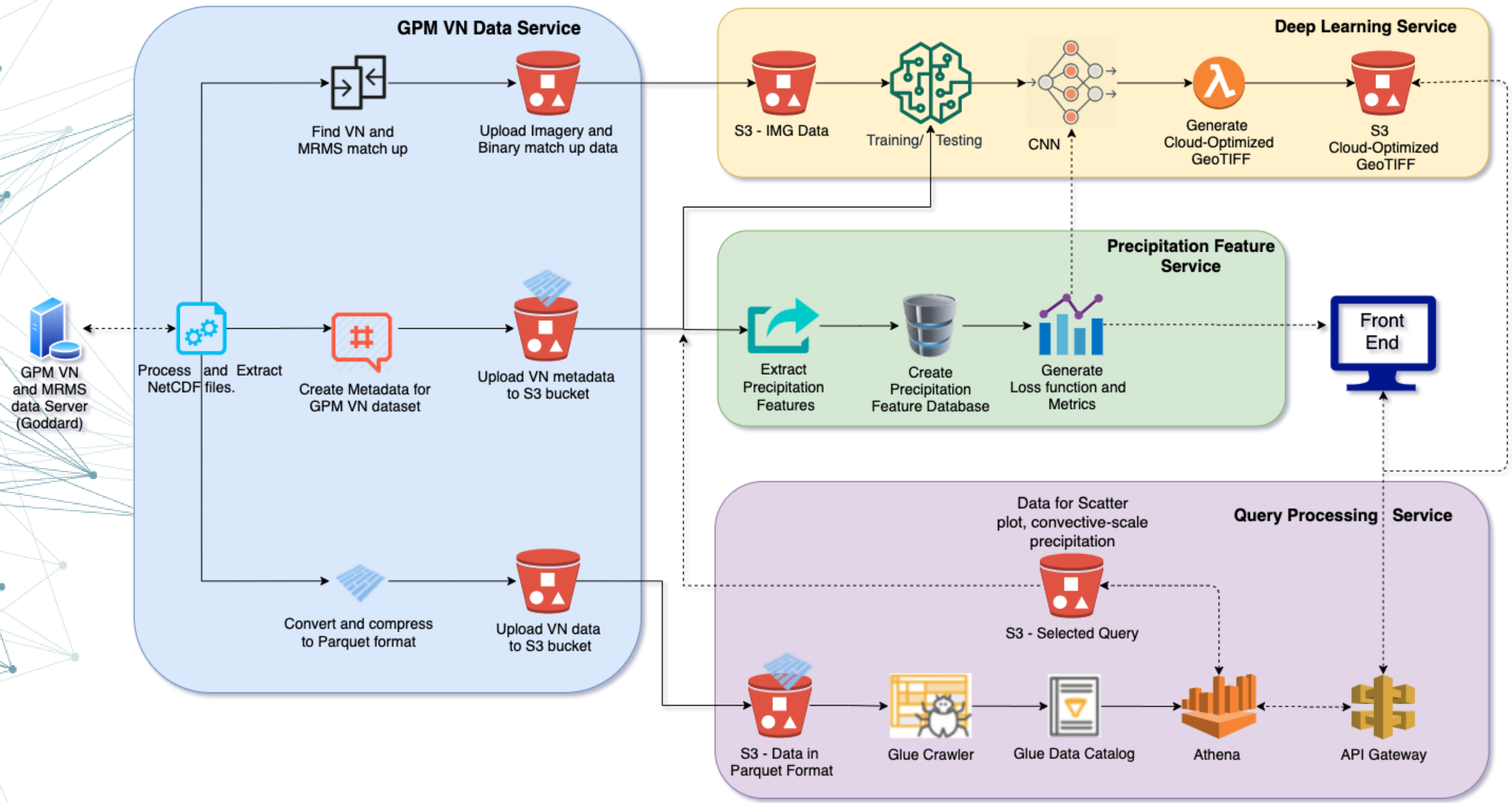
- MRMS data corresponding to the VN matchup time is extracted from the full CONUS gridded data, and a subset image is created centered on the VN radar site +/- 175km
- GPM surface precipitation from VN matchup files within the MRMS subset area are rendered as circular 5 km footprints in an image corresponding to the same area as the MRMS subset around the radar site in lat/lon .01 degree grid
- Deep learning CNN applied using MRMS data as truth to downscale GPM to MRMS resolution

Presentation Contents

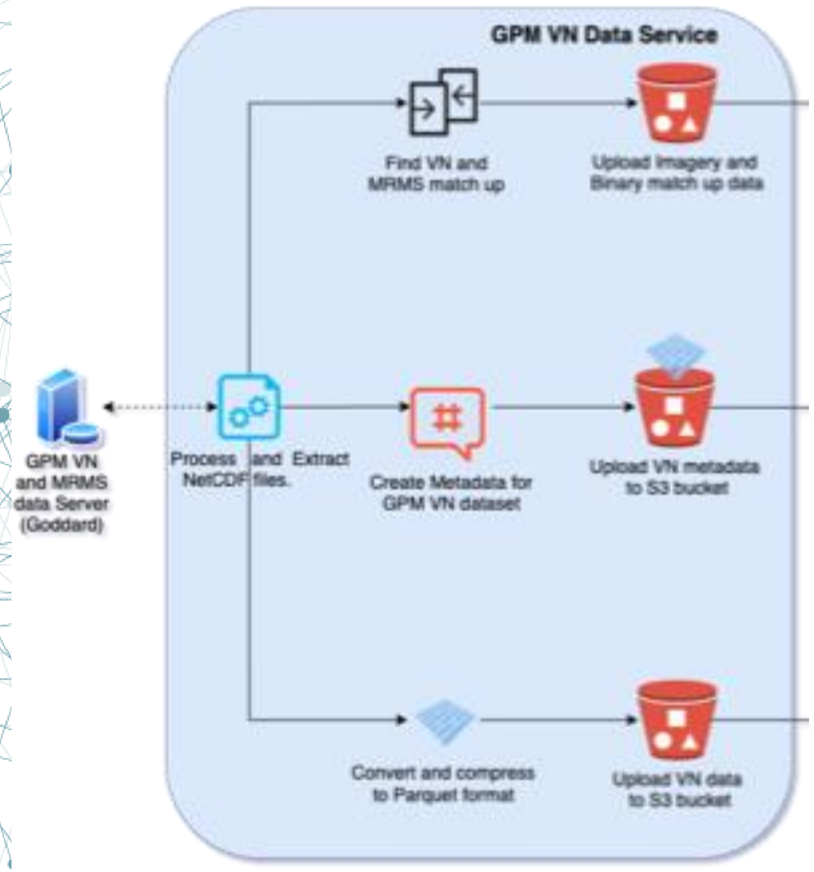
- Background and Objectives
- Technical and Science Advancements
- **Summary of Accomplishments and Future Plans**
- Publications - List of Acronyms

Summary of Accomplishments

CAPRi Architecture



Summary of Accomplishments
GPM VN Data Service

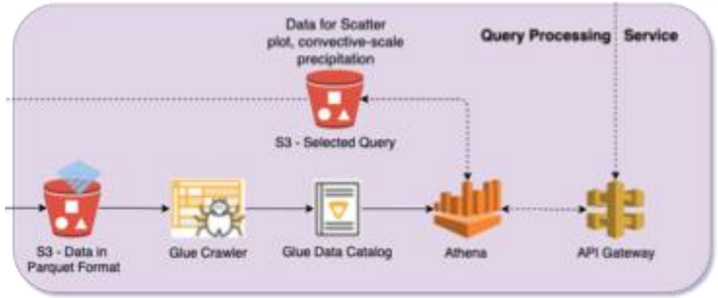


Initial Development Complete:

- Created a script to extract, process, and upload GPM VN and coincident MRMS data to CAPRi as netCDF-4 data files
- Developed a script to create both binary data and imagery of surface rain rate from GPM VN and MRMS for Deep Learning work
- Converted GPM VN 3D data from netCDF into compressed Parquet format for analysis tools
- Uploaded all the available GPM VN and MRMS matchup data and imagery to AWS S3 buckets
- Ran Unit Tests on GPN VN Data Service
- Retrieved additional data using CAPRi services
- Created an Athena partition for optimizing performance

Summary of Accomplishments
Query Processing Service

- Completed a comparative study on data formats based on storage size, retrieval speed, and best compression ratio. (JSON, CSV, and Parquet)
- Setup a Glue crawler to read compressed Parquet data from AWS S3 buckets then catalog and store in AWS Athena
- Implemented security controls on the API
- Created queries to request data in AWS Athena
- Created an API to query data from AWS Athena
- Created support to utilize the API for front end and as stand-alone
- Wrapped the project using Terraform Framework to easily provision services
- Resolved cross-origin resource sharing (CORS) issues for the API
- Investigated options to resolve the latency in API calls
- Implemented paging options and download options
- Created a Python Class for the Science team to utilize the stand-alone API



Summary of Accomplishments
CAPRi Core API Definition

- The CAPRi Core API provides endpoints for VN matched precipitation data. Data are first ingested in S3 bucket and then are indexed with the corresponding key into AWS Athena. AWS Athena stores the data in columnar format and can be queried with SQL queries. A Python class has been developed that allows a programmer to easily set up and perform queries of VN data in Athena without having to deal with complicated URL construction and SQL query commands. Query results can be streamed directly to a python data structure or downloaded to a file for later analysis.

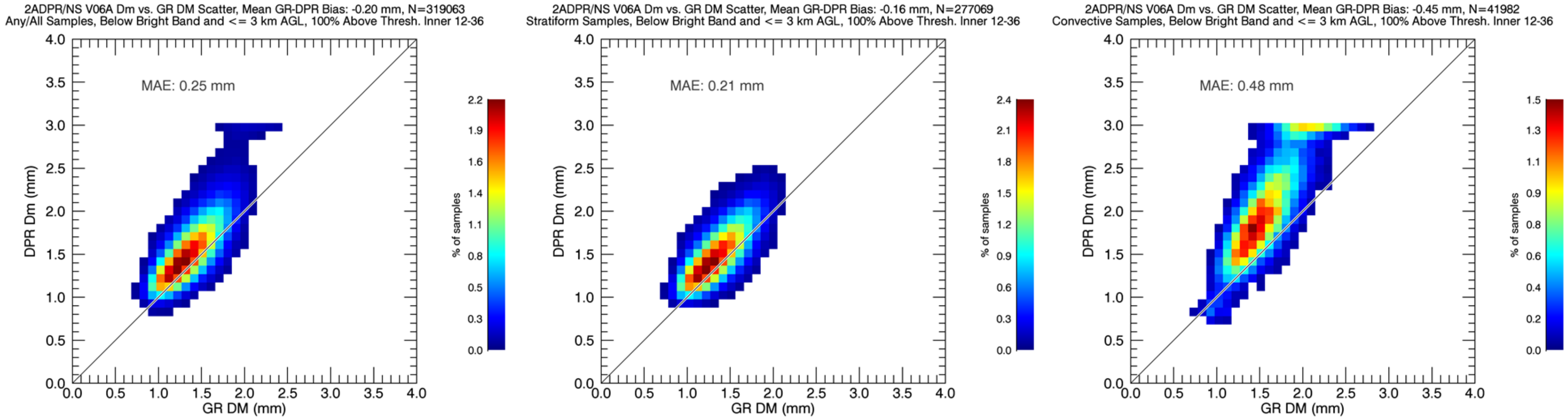
- Co-I Gatlin is evaluating a prototype of the code and has provided useful feedback and suggestions for improving performance and functionality.

Summary of Accomplishments

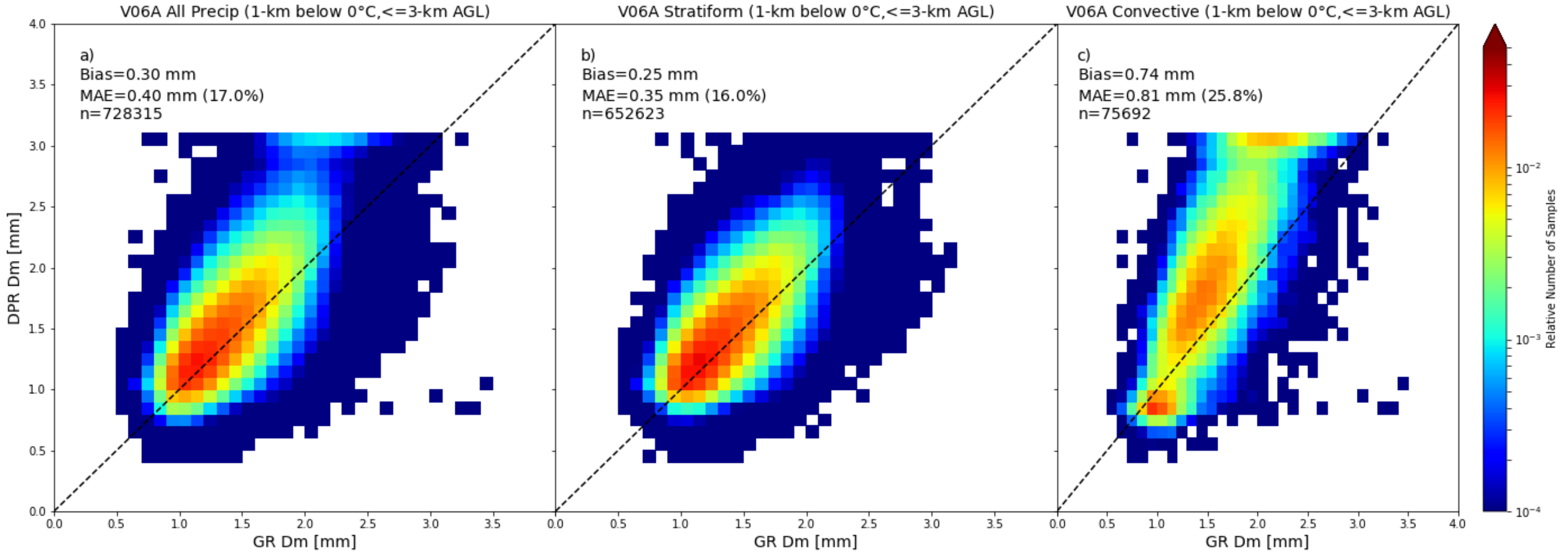
Use Case: Evaluation of satellite retrieval

Facilitates a 48x faster analysis of VN dataset

Traditional
File-based access
IDL
(35min)

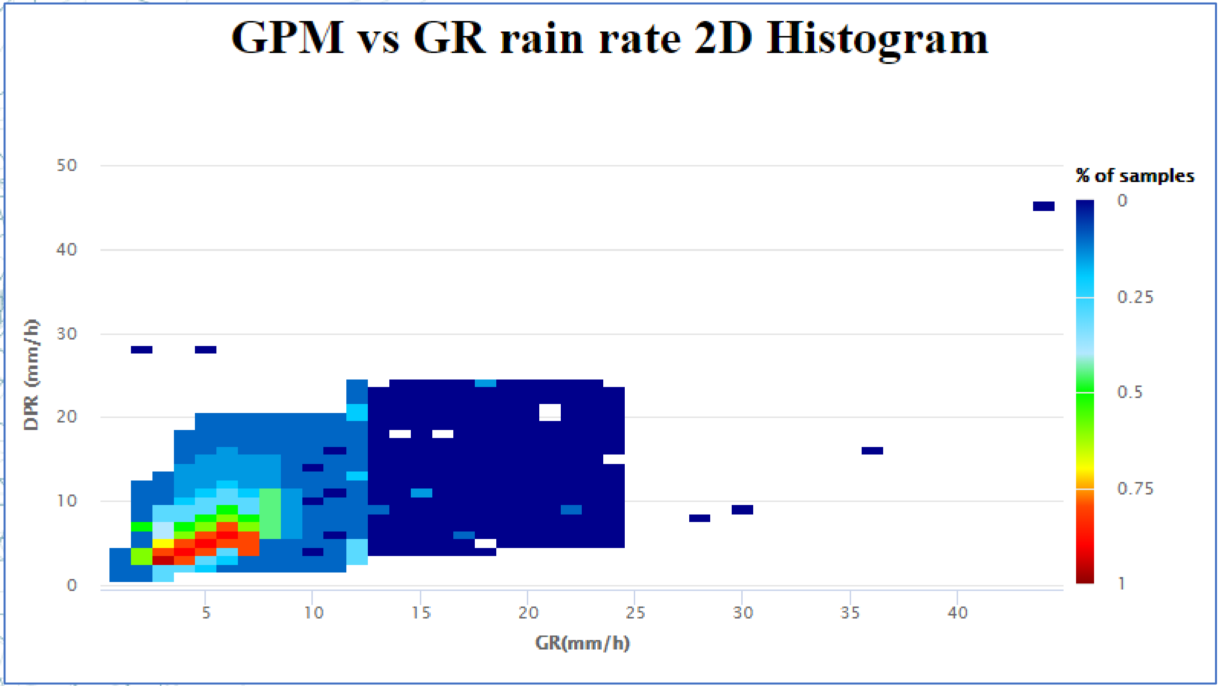


CaPRi
Athena
Python
(*<43s*)



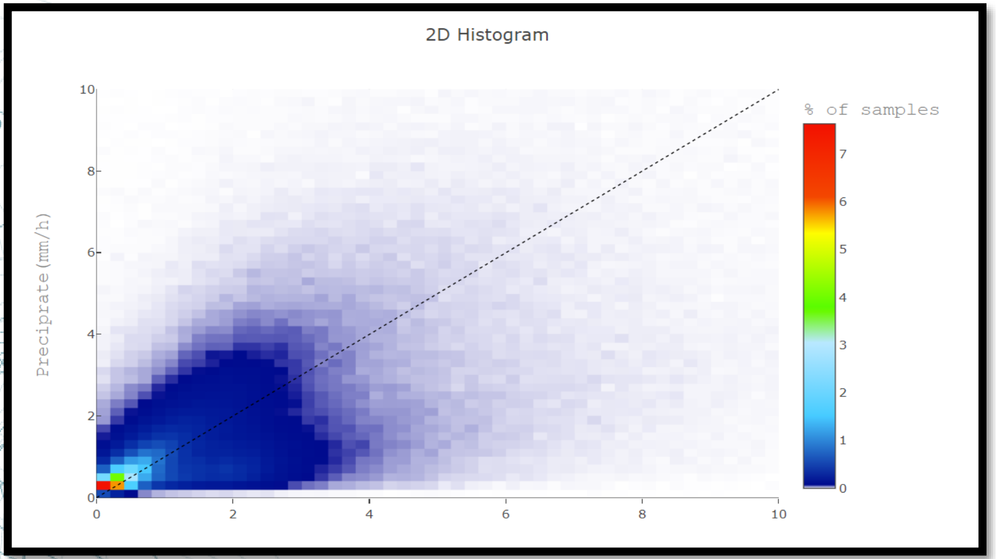
Summary of Accomplishments

Front End

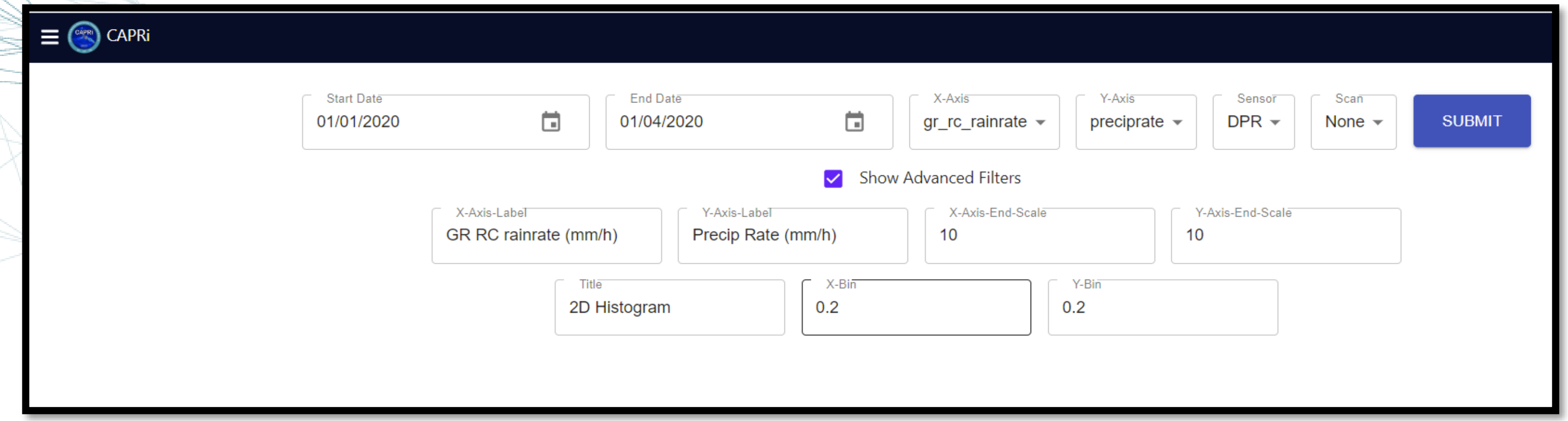


- Investigated various charts and renderings to create a meaningful representation of the datasets to include Histogram, Heat map, Scatter plot, and Line graph
- Developed an REACT webpage to interact with AWS API Gateway to fetch sample data for charts
- Conducted a case study to place the components across the webpage and developing the wireframes
- Selected CesiumJS for dataset visualization and investigating deck.gl for additional capabilities (i.e., large dataset visualization)

Summary of Accomplishments
Front End



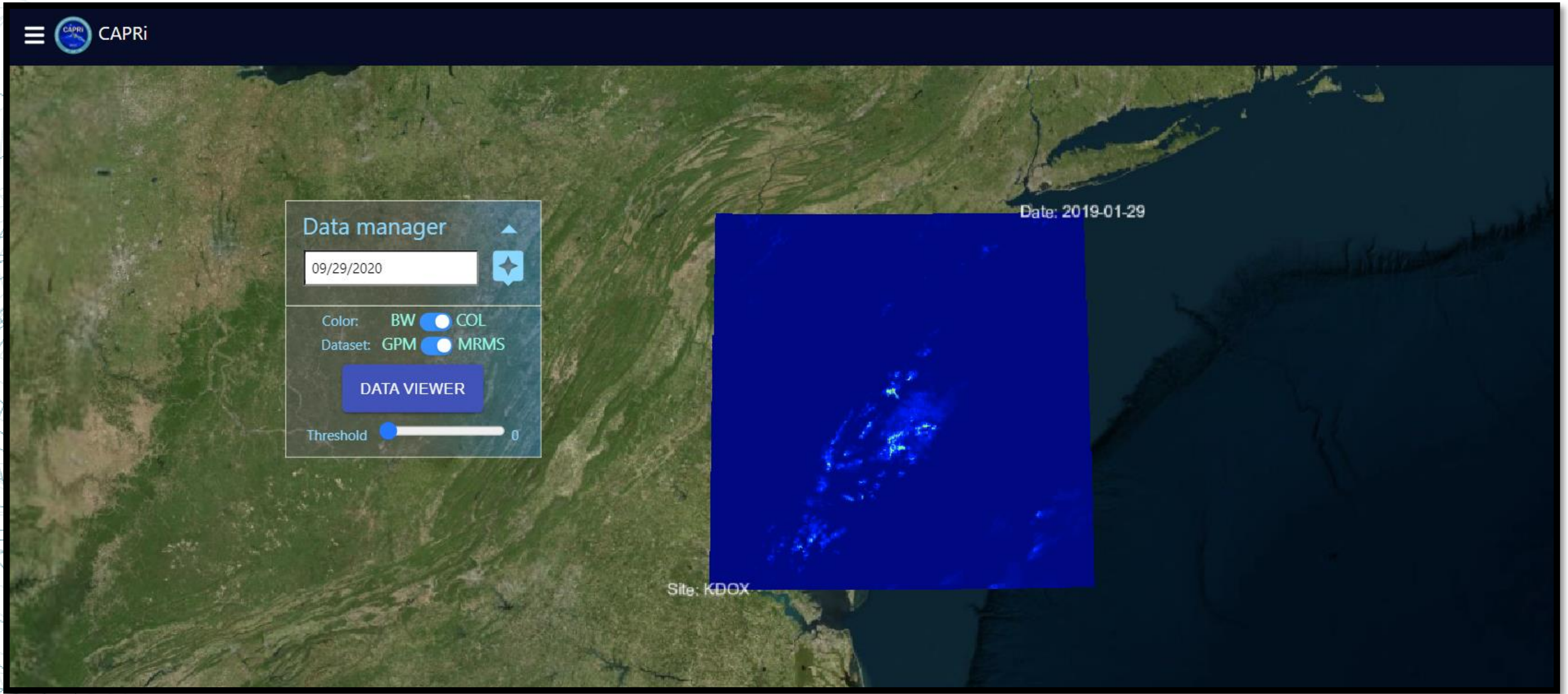
- Adjusted the chart rendering capabilities based on advice from Co-I Gatlin and Co-I Berendes such as implemented 2D histogram plot which automatically calculates the bin intervals and color codes the number of samples.
- Added the sensor and scan type filters.
- Added advanced filters edit customizing the scale of X or Y-axis and the binning interval sizes. User can now even edit the title, X or Y axis labels etc.



Summary of Accomplishments
Front End

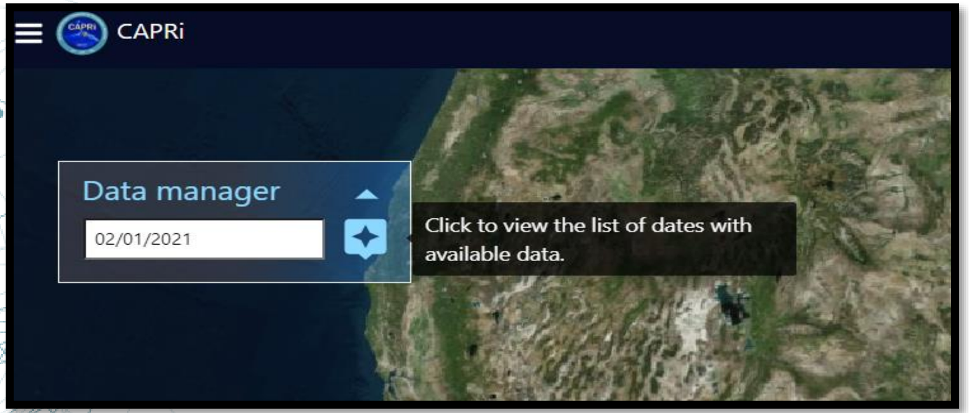
- Designed an API for accessing KML images containing the date, radar site,   GPM orbit, PPS version, GPM sensor, GPM scan, and VN version number.
- Utilized the Cesium application to develop a Front-End mapping component.
- Added the KML images dynamically to the map according to the selected filters.
- Finished adding a transparency slider for the image display flipping MRMS and GPM sites.
- Added a results dialog box called “Data Viewer” where image locations for the selected date are visible.
- Added toggle switches for Color to Black and for selecting datasets (MRMS or GPM).

Summary of Accomplishments
Front End

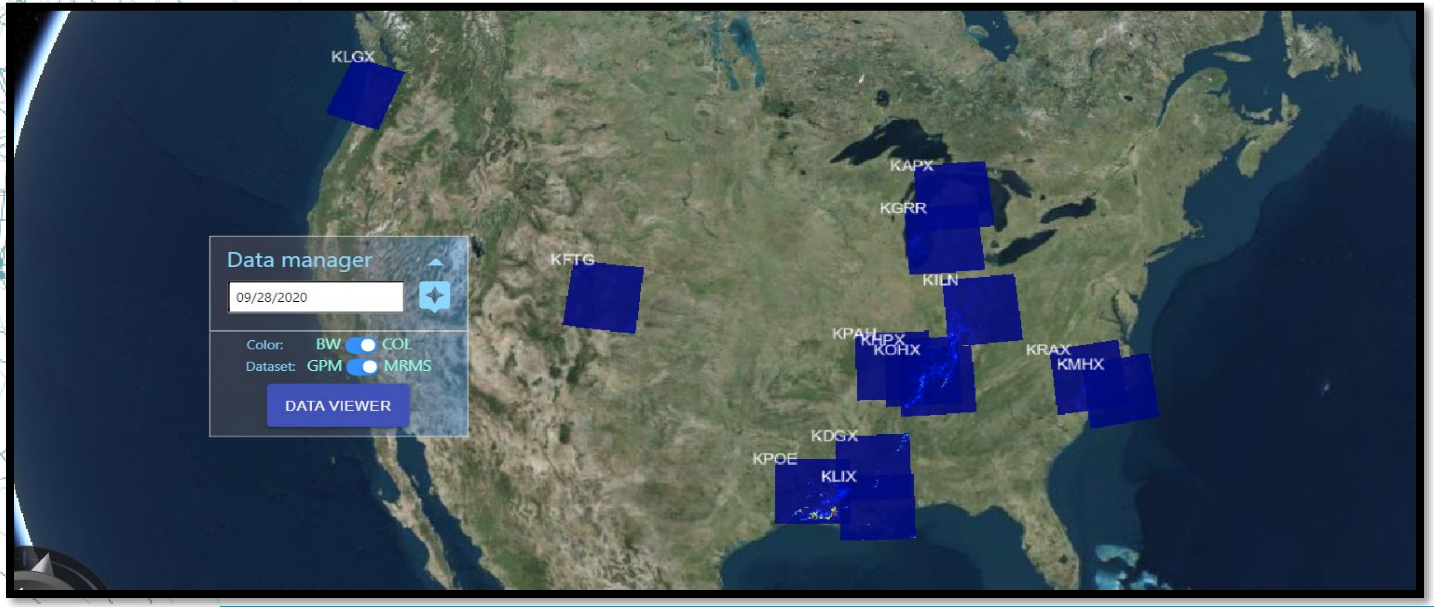


Example of the Front-End mapping capability

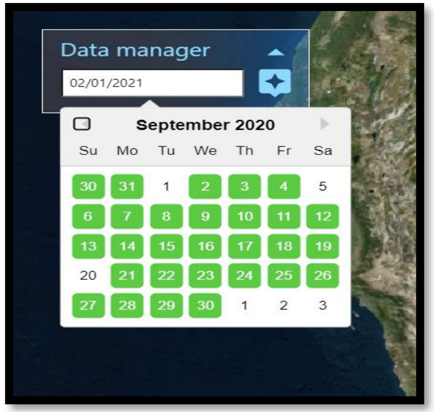
Summary of Accomplishments
Front End



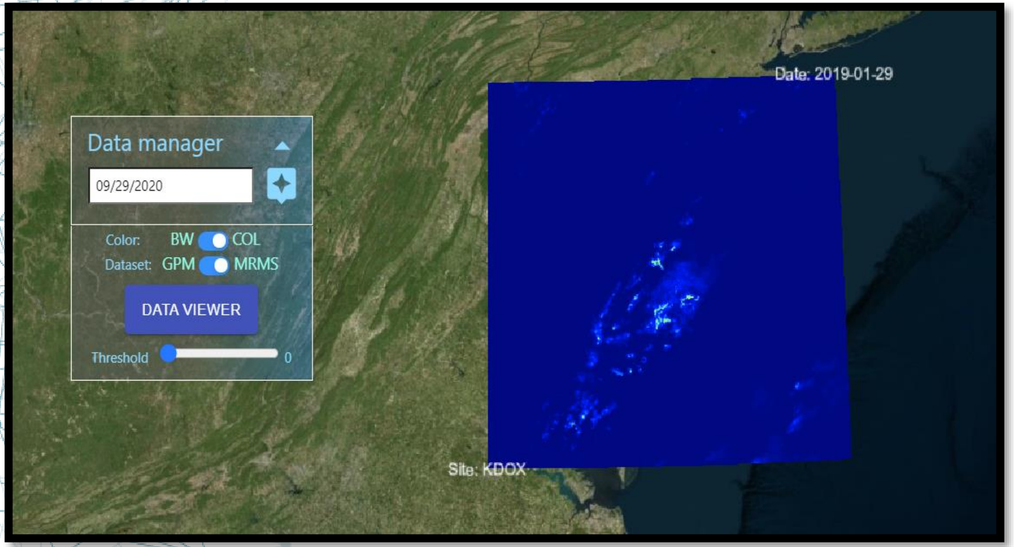
- Updated the Data manager toolbar with a date picker and an assistant icon(with a tooltip). Clicking on the assistant gives a list of all the dates available with data.
- Updated the Data manager toolbox to be movable.
- Selecting a date auto-submits and plots the images for all sites. Added a toggle to switch from MRMS to GPM and from Color to Black & White images.



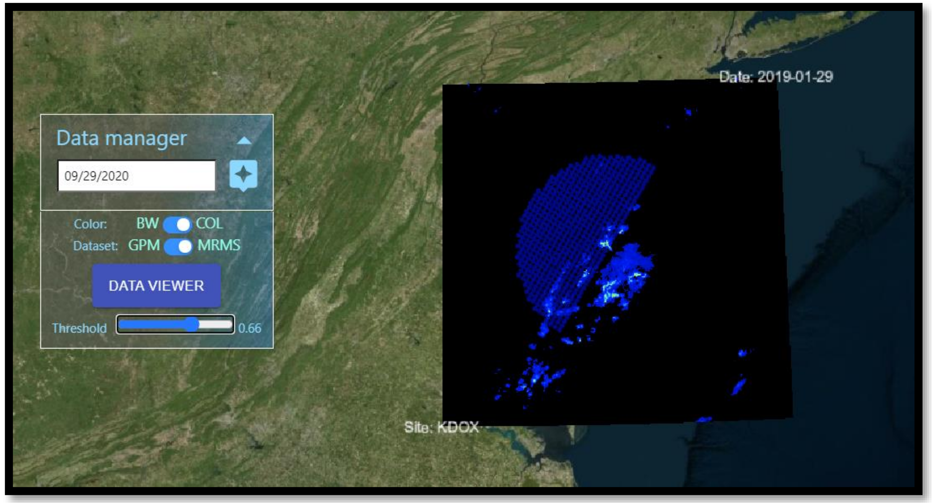
- Added a highlighter in the date picker for dates with available data.



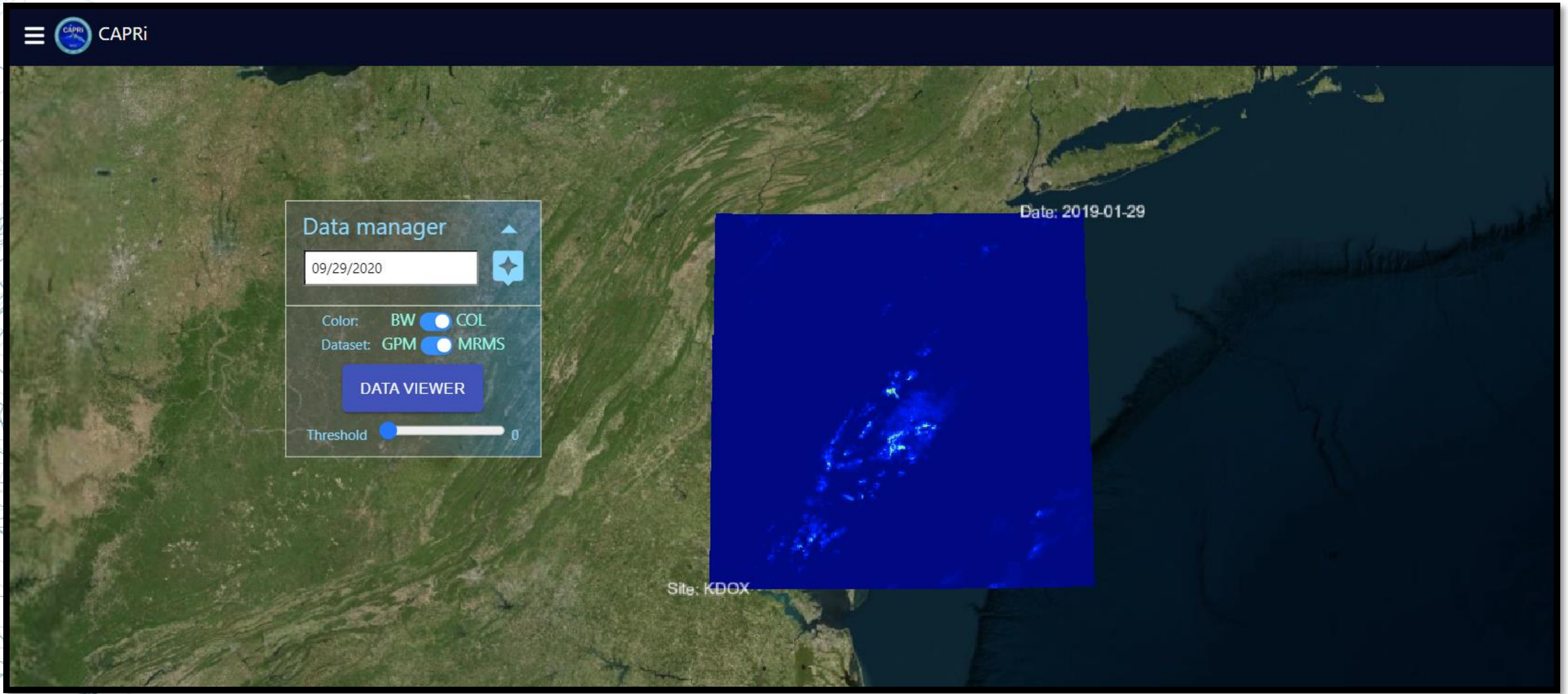
Summary of Accomplishments
Front End



- Added labels to images like site names (when all images for a particular date are viewed at once) and date and site name (for single image view).
- “Data Viewer” has list of sites. Clicking on each row opens the corresponding image and the transparency slider.
- Sliding the Threshold value lets you extract the hotspots and compare with its corresponding GPM image.
- Adding a color scale for images is underway.



Summary of Accomplishments
Front End



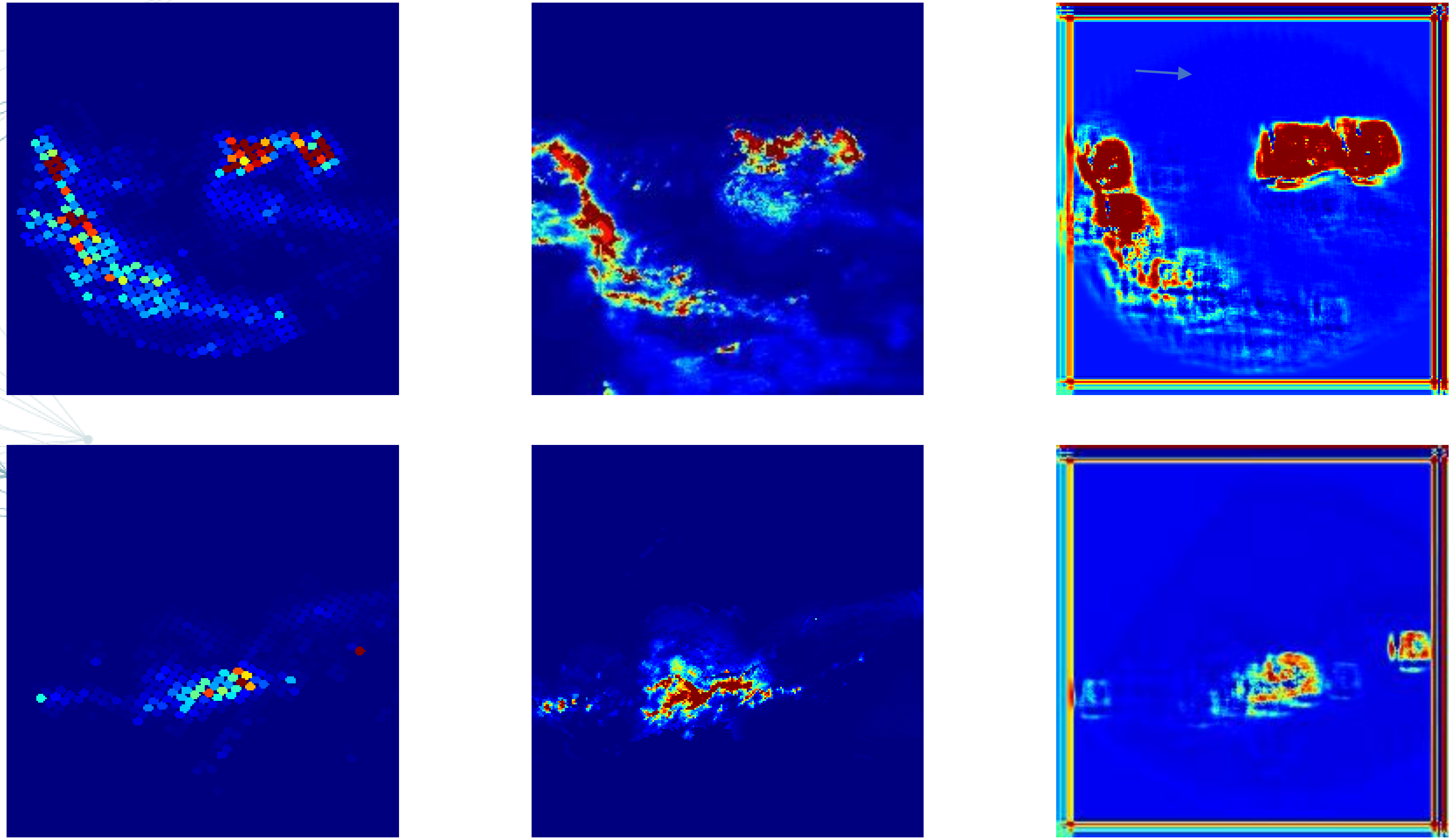
- Example of the front end mapping capability

Summary of Accomplishments
Deep Learning

- Two models are being developed and tested: Unet and Generative Adversarial Network (GAN)
- An Unet learns the features of the input GPM DPR data through a series of convolutional and activation layers and then generates the output from these features. The output is then evaluated against the corresponding MRMS sample.
- GANs consist of two constituent neural networks: Discriminator Network and Generator Network
- The generator network is trained to generate enhanced GPM DPR data using corresponding MRMS data as a target.
- The discriminator network is trained to distinguish the generator's output and real MRMS data.
- The zero-sum gain between these two networks results in the generator network learning how to generate data that closely resembles MRMS data.
- Due to the presence of no data gaps in the GPM DPR data, a second band is added to the rain-rate matrix.
- This band consists of binary flags at each index indicating the presence (or absence) of an actual measurement value.
- This is intended to help the network learn to approximate values in the areas for which there are no measurements in the GPM DPR data matrices.

Summary of Accomplishments
Deep Learning – Super Resolution GAN (SRGAN)

GPM DPR | MRMS | SRGAN



Summary of Accomplishments

Deep Learning – Super Resolution GAN (SRGAN)

GPM VN geomatch files

Gridded MRMS sfc precip

VN data fusion with MRMS

MRMS grid merged to GPM VN gemoatch 3D volume product netcdf

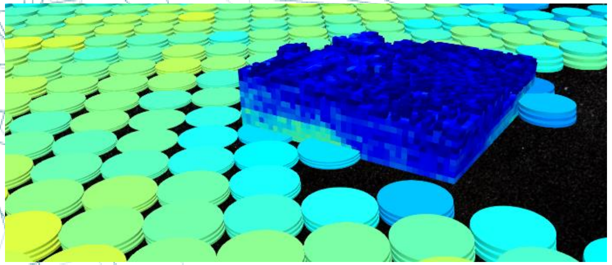
3D VN volume metadata and statistics (json) ingested into AWS Athena database (GPM and GR, MRMS could be added))

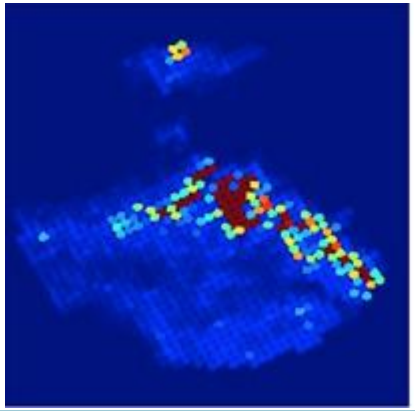
GPM VN and MRMS imagery/binary rendered to MRMS lat/lon grid

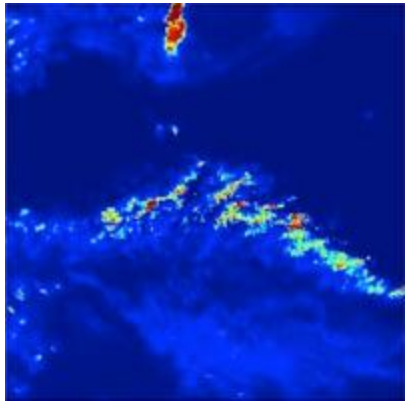
Gridded binary files for MRMS and GPM and MRMS sfc precip

GPM lat/lon image surface precip rate

MRMS lat/lon image surface precip rate







Future Plans

- Finalize the development of the Front-End and GPM VN Data Services
- Finalize decisions on data retrieval methods from the Query Processing Service
- Complete the design and Implement the Deep Learning Service
- Develop a module to generate Training and Testing Data
- Start developing the Precipitation Feature Service with Test Data

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Publications

Presentations

- Conducted an on-line presentation at the 2020 ESTO Technology Forum on 24 June 2020.
- PI (Beck) presented the project at the 2020 Annual American Geophysical Union (AGU) Conference.
- Co-I Gatlin presented some results from using the GPM VN API at the 2020 Annual American Geophysical Union (AGU) Conference.

Acronyms

3D	Three Dimensional
API	Application Programming Interface
CNN	Convolutional Neural Network
CORS	Cross-origin resource sharing
DPR	Dual-frequency Precipitation Radar
GPM-VN	Global Precipitation Measurement-Validation Network
IMERG	Integrated Multi-satellite Retrievals for GPM
LIS	Lightning Imaging Sensor
MRMS	Multiple Radar Multiple Sensor

Valid time-series analyses of satellite data to obtain statistical inference about spatiotemporal trends at global scales

Anthony R. Ives (PI, University of Wisconsin-Madison)
Jun Zhu (Co-I, University of Wisconsin-Madison)
Volker C. Radeloff (Co-I, University of Wisconsin-Madison)
Fangfang Wang (Co-I, Worcester Polytechnic Institute)

AIST-18-0020 Annual Technical Review
5 February, 2021

Analyzing remotely sensed time-series data

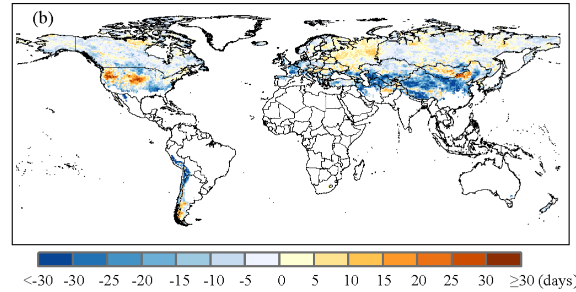
PI: Anthony Ives

Objective

Develop new statistical tools to analyze large, remotely sensed datasets that will give statistical rigor to conclusions about patterns of past change and statistical confidence to forecasts of future change.

Our goal is to develop methods for valid statistical tests of hypotheses involving changes through time at the spatial resolution of many remote sensing products (containing potentially millions of non-independent time series).

The methods will be general enough to address diverse types of hypotheses, such as whether changes in the duration of snow cover vary with latitude, and whether land-cover classes (e.g., cropland, deciduous forest) differ in their rates of greening.



Example: Changes in the number of days of frozen ground without snow cover per year based on AMSR-E and JAXA satellite data for 1982-2014. This dataset consists of >6,000,000 correlated time series that challenges hypothesis tests by current statistical methods.

Approach

We will build:

1. statistical algorithms for analyzing large spatiotemporal datasets. Our strategy is to separate the task of analyzing time series from the task of analyzing spatial structure, so that computation burden scales linearly with data size. Our approaches to this problem will be general and can be applied to many types of models of temporal change.
2. a software package that applies the algorithms.

We will test these algorithms with AVHRR/GIMMS3g, MODIS, AMSR-E, JAXA/JASMES, and Landsat data

Co-Is: Volker Radeloff, Fangfang Wang, Jun Zhu

Key Milestones

Year 1

5/20 TRL3 for Partitioned Autoregressive Time Series (PARTS)
8/20 TRL4 for PARTS
11/20 TRL5 for PARTS
2/21 TRL3 Panel regression (PR)

Year 2

4/21 TRL4 for PR
8/21 TRL5 PR
11/21 Applications for PARTS and PR
2/22 Completed software (R package)

$TRL_{in} = 2$

Presentation Contents

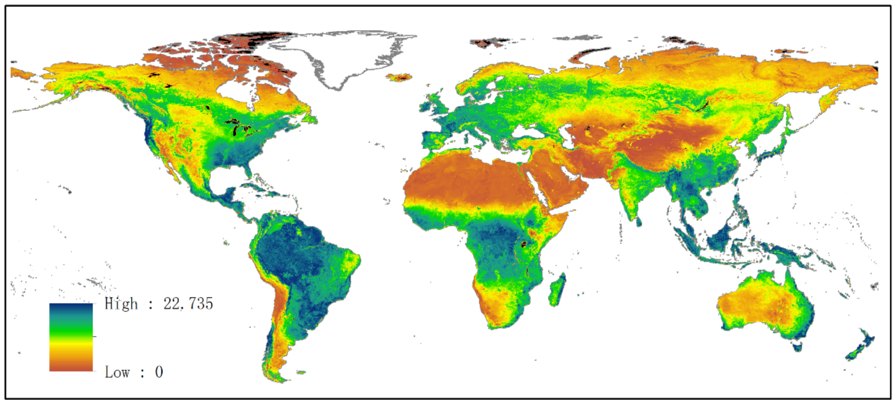
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Earth Science Technology Office

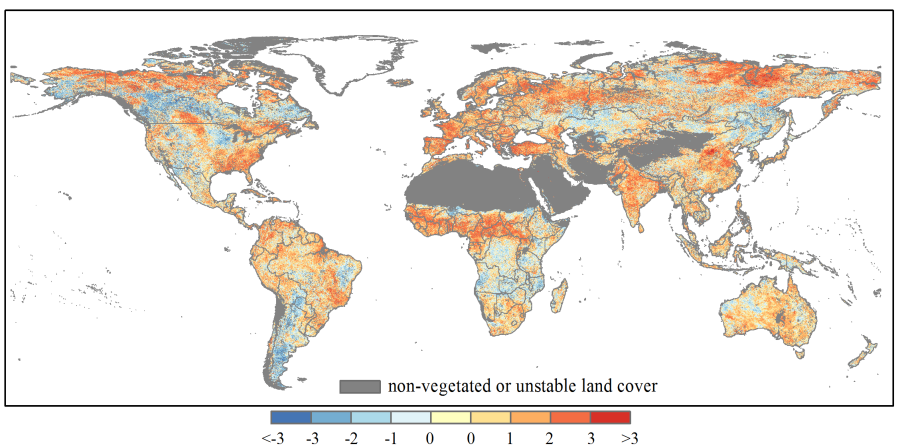
Background and Objectives

Remote sensing provides vast amounts of data.

Making the most of them requires appropriate statistical models.



Mean greenness (NDVI), 2013

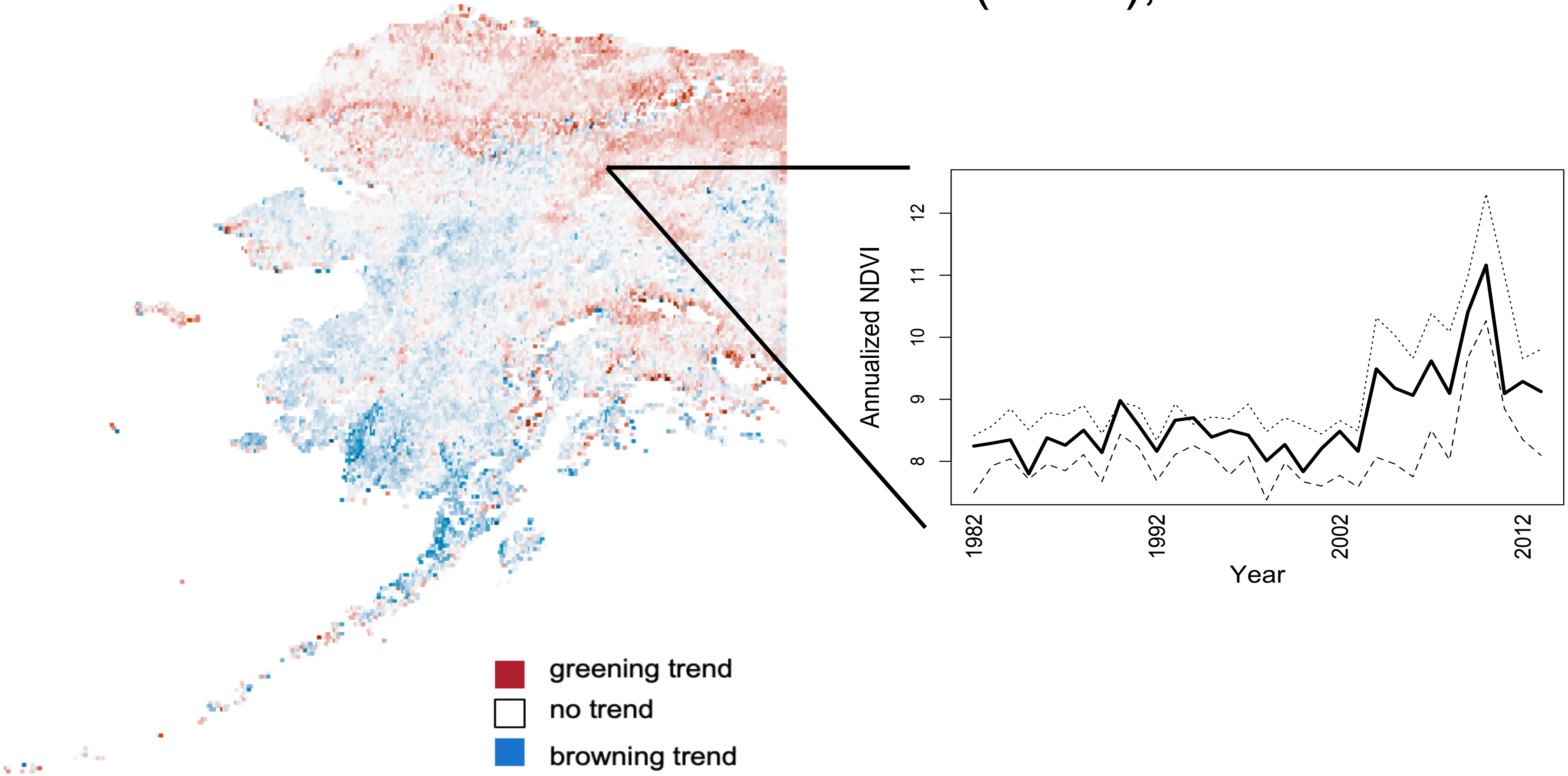


Change in greenness (NDVI), 1982-2013

Background and Objectives

Alaska west of –141°

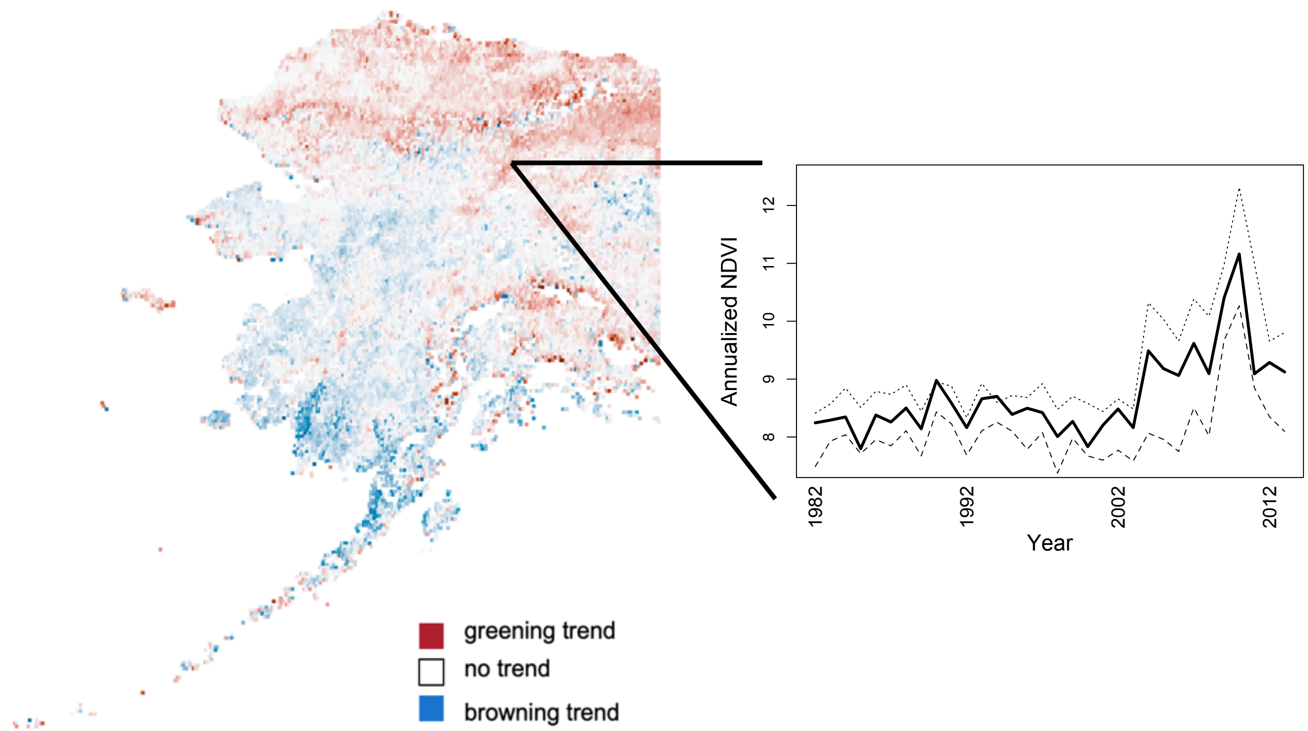
Change in greenness (NDVI), 1982-2013



Background and Objectives

Hypotheses involve broad-scale (not within-pixel) patterns

Statistical models need to account for temporal and spatial autocorrelation in the "unexplained" variation.



Background and Objectives

Hypotheses involve broad-scale (not within-pixel) patterns

Statistical models need to account for temporal and spatial autocorrelation in the “unexplained” variation.

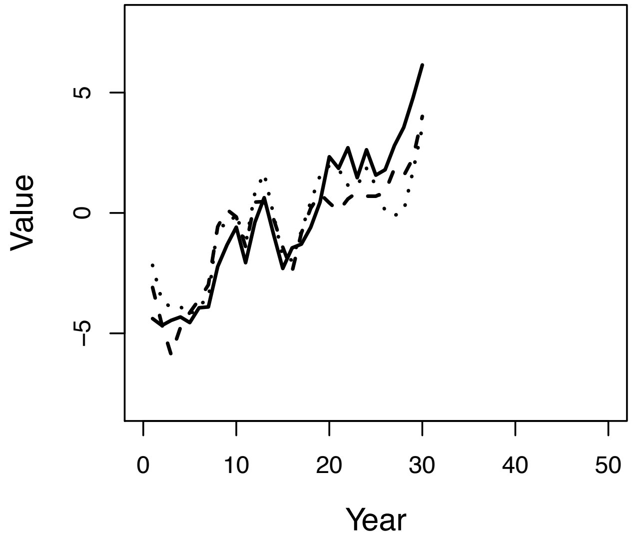
<u>Example illustrations with simulations</u>

1. Time trends vs. temporal autocorrelation

2. Spatial drivers vs. spatial autocorrelation

3. Revealing true drivers underlying observed patterns

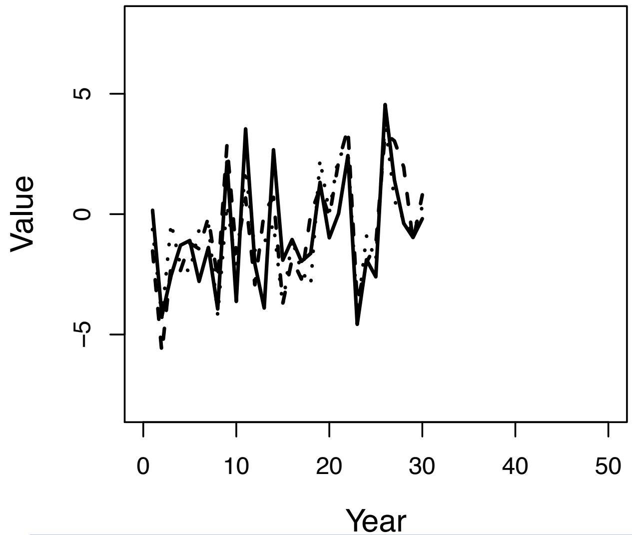
Background and Objectives



A

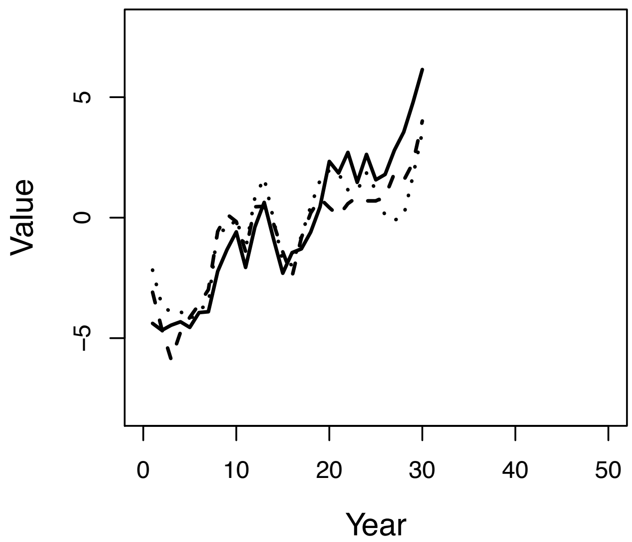
1. Time trends vs. temporal autocorrelation

Which time series has the greatest change, A or B?



B

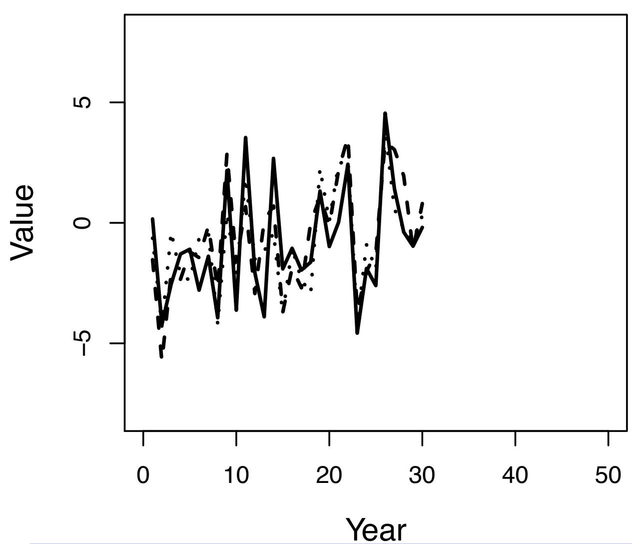
Background and Objectives



A

$$x(t) = bx(t\text{-}1) + \varepsilon(t)$$

Autocorrelated process

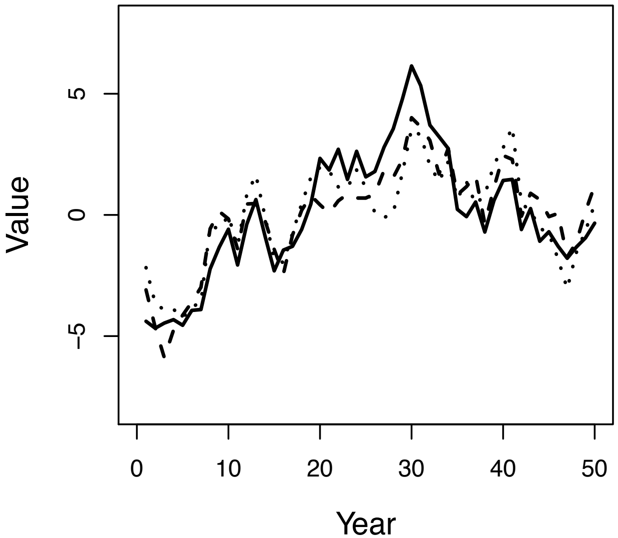


B

$$x(t) = a + ct + \varepsilon(t)$$

Time trend

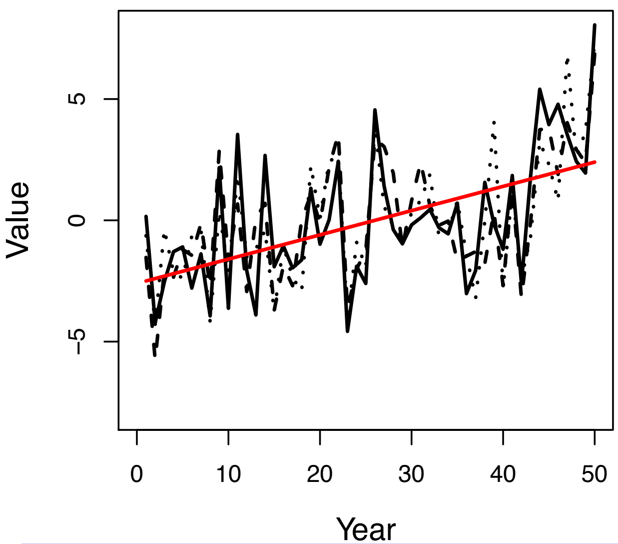
Background and Objectives



A

$$x(t) = bx(t\text{-}1) + \varepsilon(t)$$

Autocorrelated process

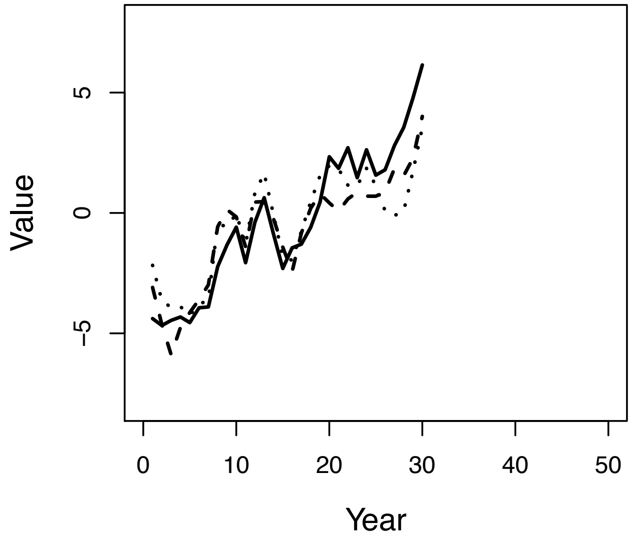


B

$$x(t) = a + ct + \varepsilon(t)$$

Time trend

Background and Objectives

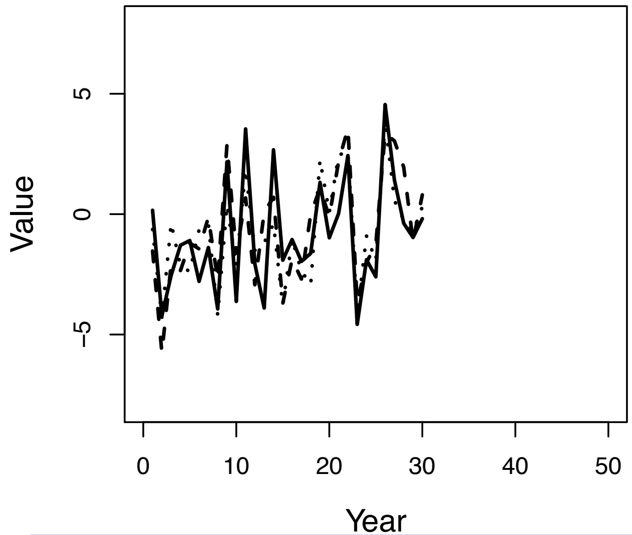


A

Which time series has the greatest change?

The differences between the first and last points are greater in A than B

Focus on one outcome of a statistical process

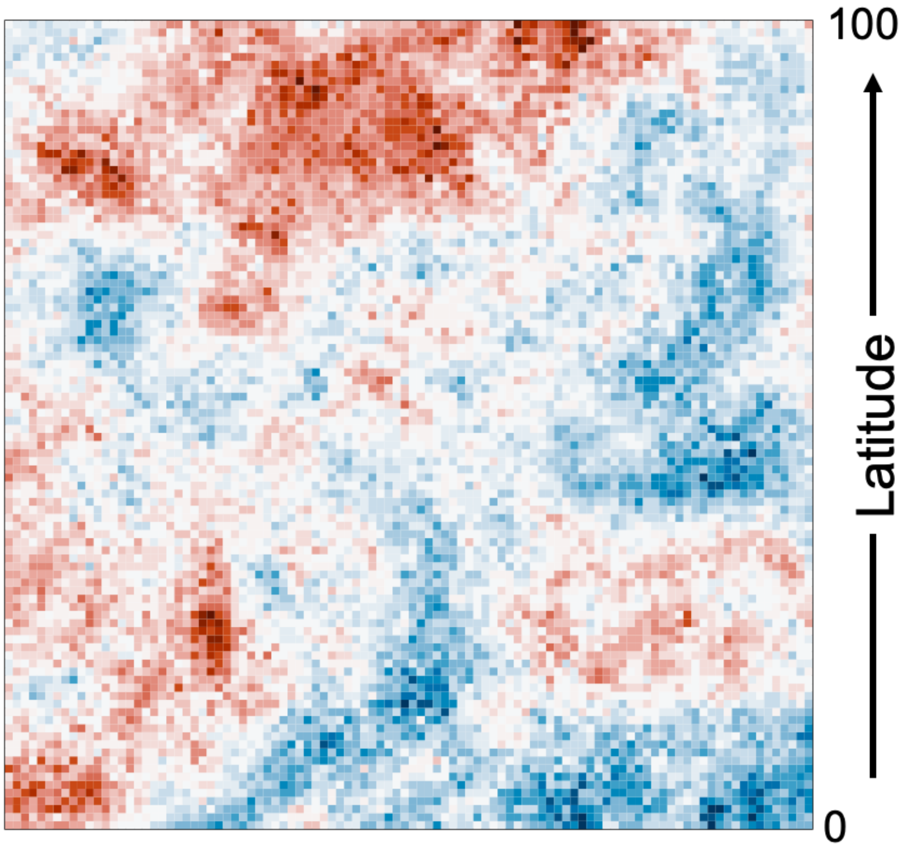


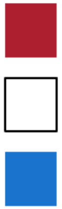
B

Statistically, the expected value increases in B but not A

Focus on the process

Background and Objectives

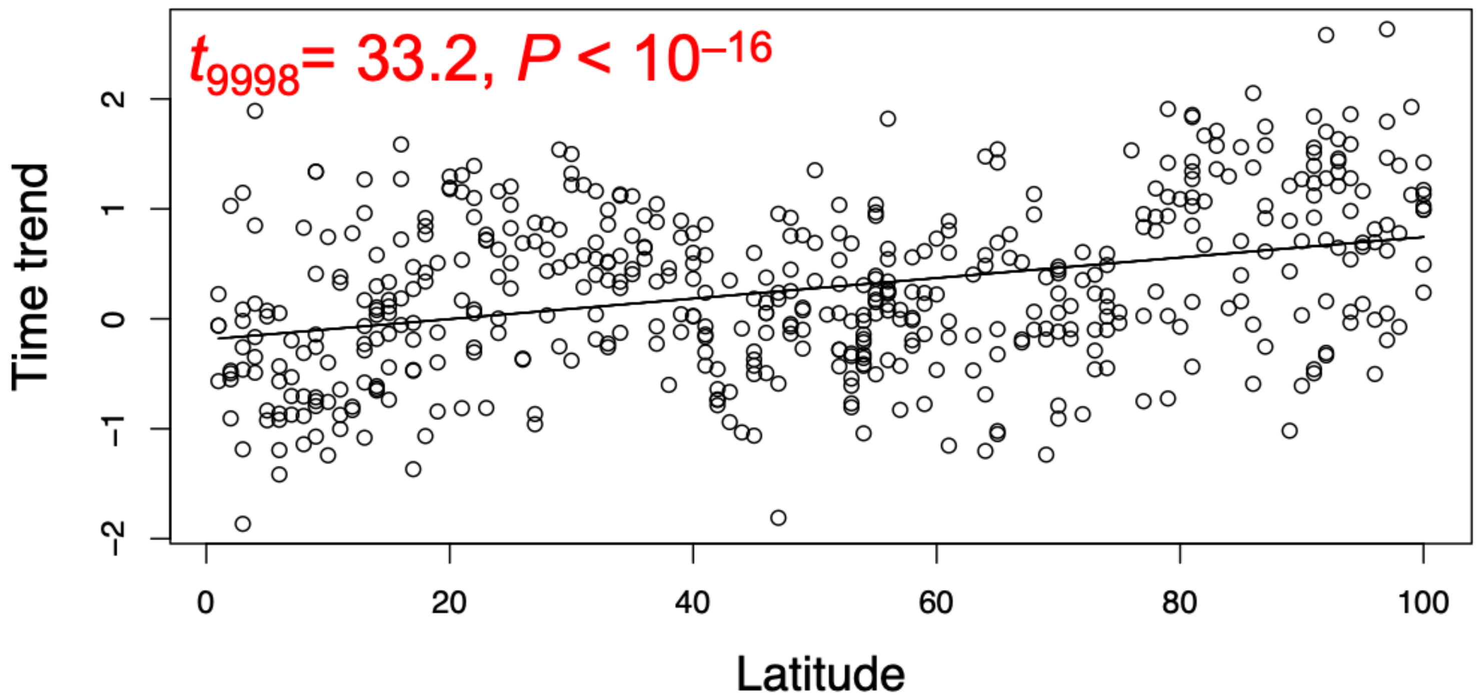




2. Spatial drivers vs. spatial autocorrelation

Is there greater greening at higher latitude?

Background and Objectives



showing trends for 500 of 10,000 pixels

Background and Objectives

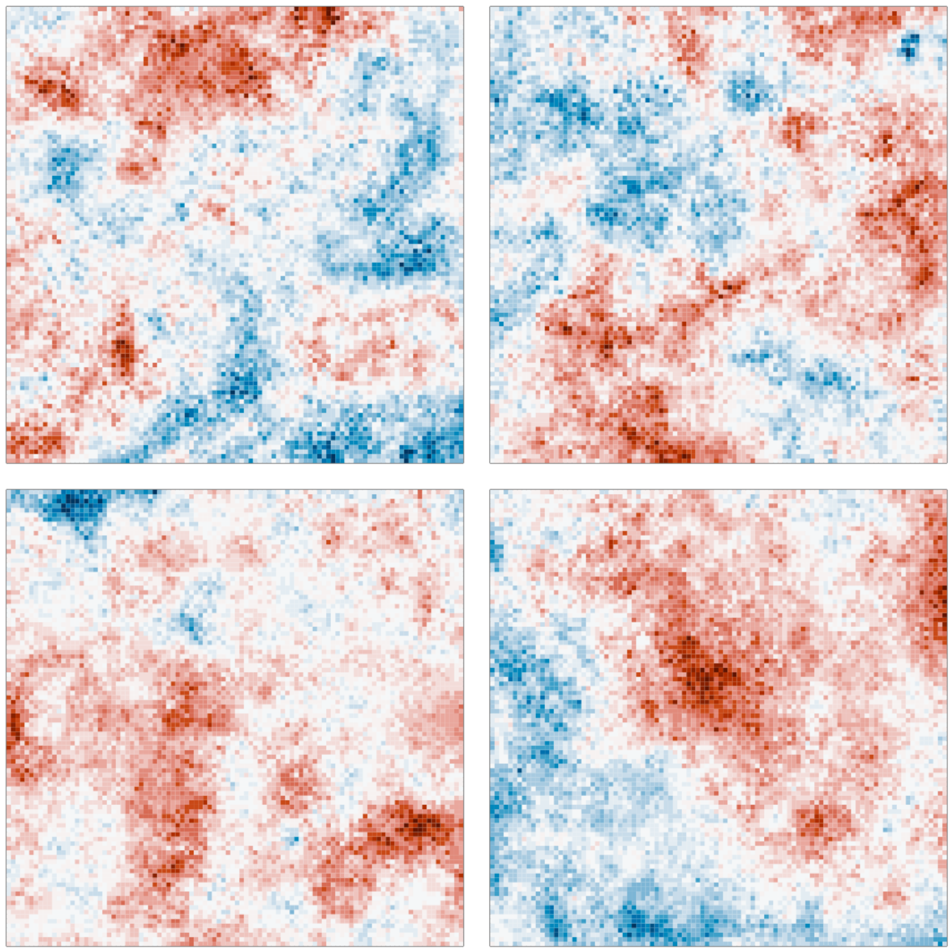
Is there greater greening at higher latitude? NO

Simulation 1

Simulation 3

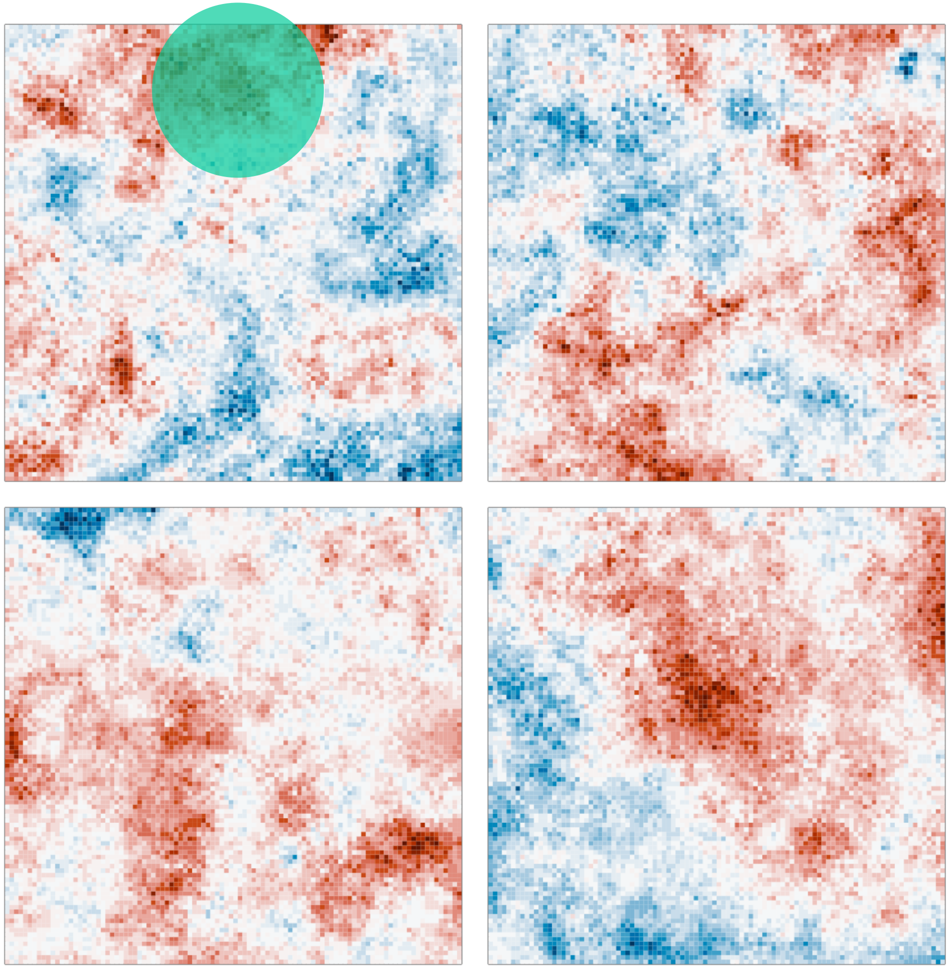
Simulation 2

Simulation 4



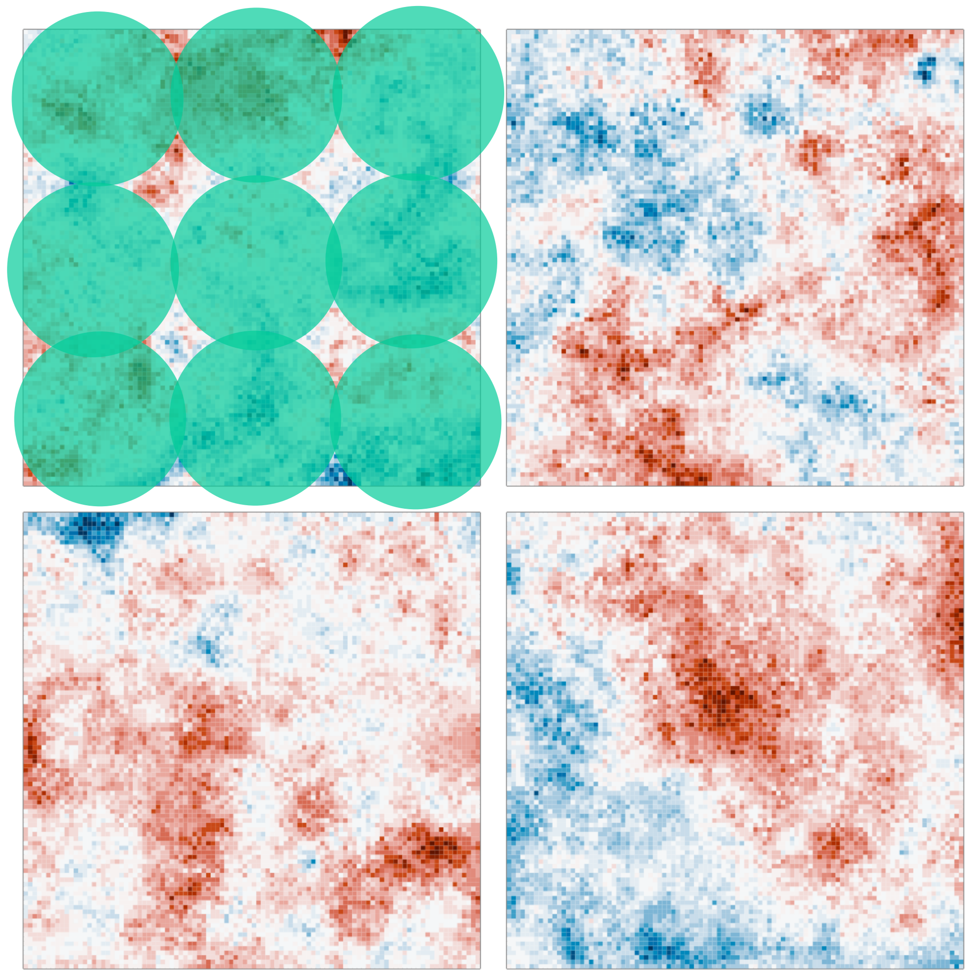
Background and Objectives

What you are seeing is spatial autocorrelation



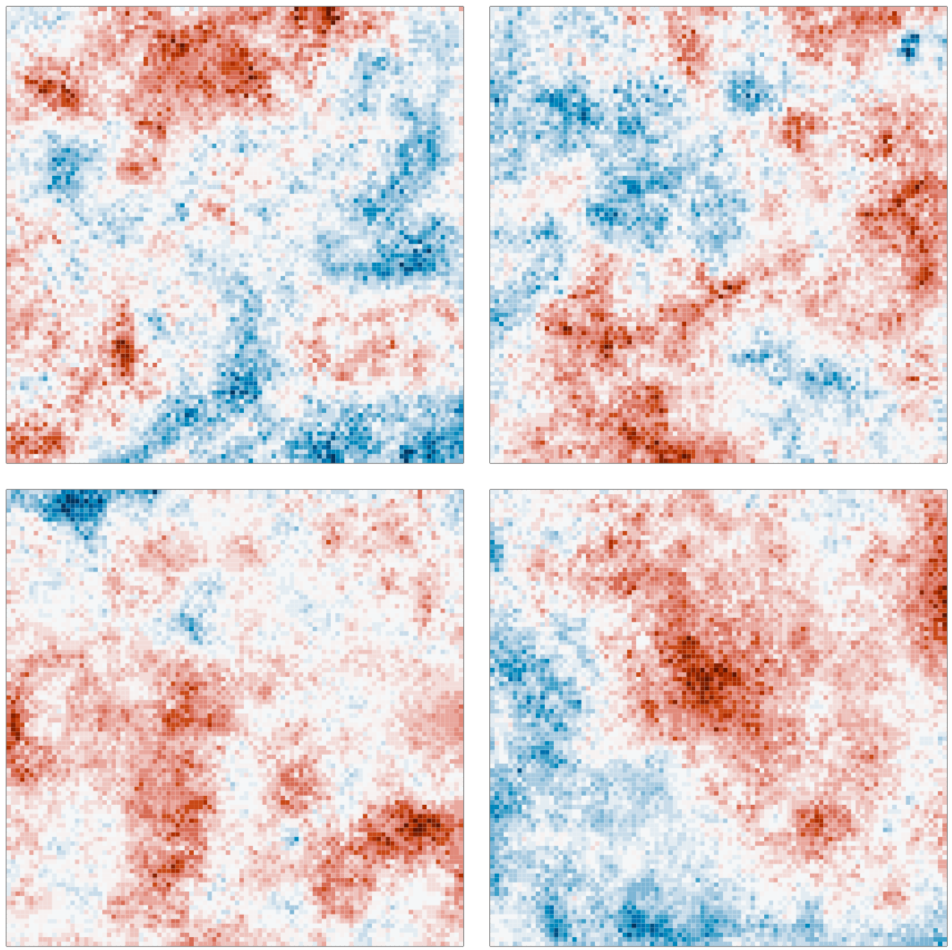
Background and Objectives

What you are seeing is spatial autocorrelation



Background and Objectives

Partitioned Autoregressive Time-Series (PARTS) analysis



P = 0.44

P = 0.85

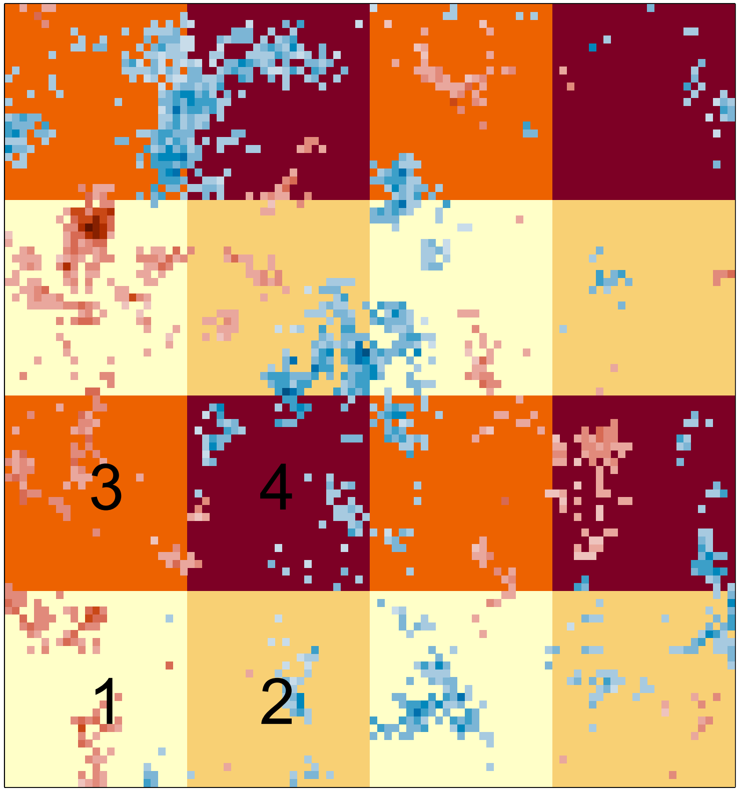
P = 0.17

P = 0.16

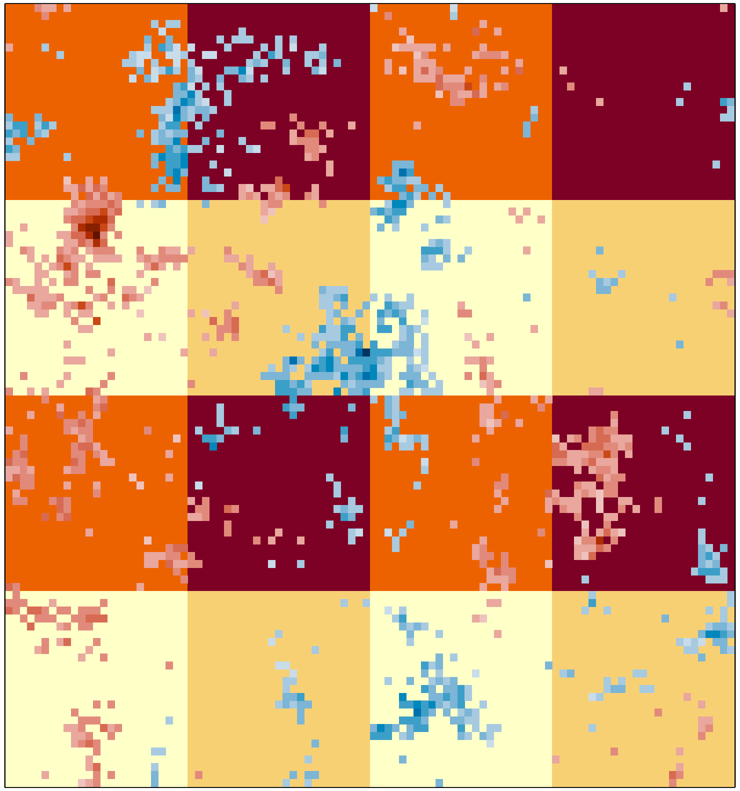
Background and Objectives

3. Revealing true drivers underlying observed patterns

A: 1 = 2 = 3 = 4

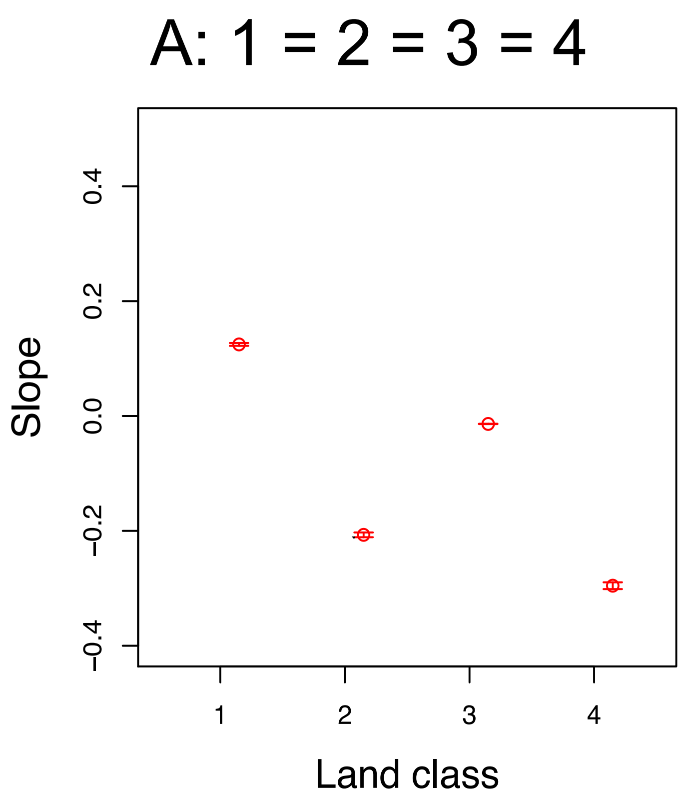


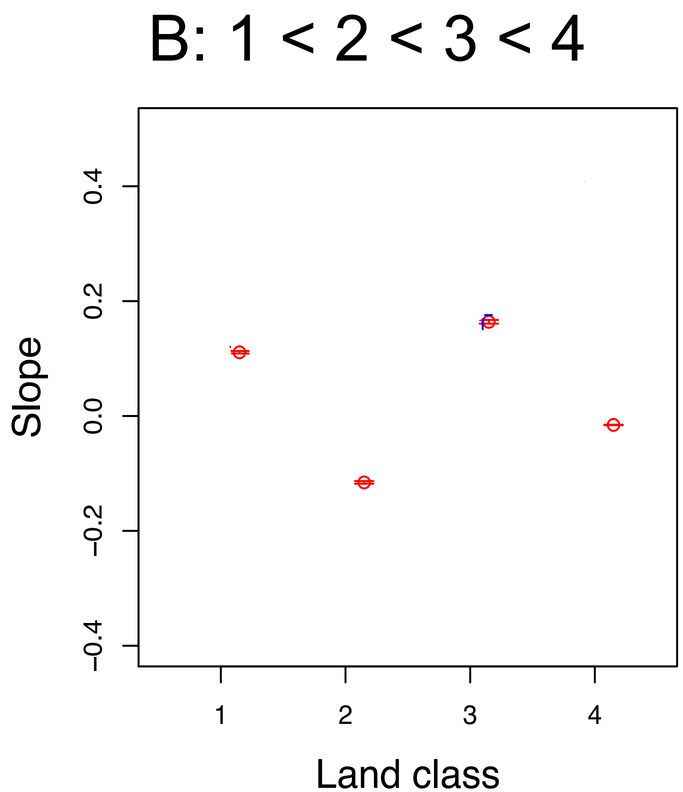
B: 1 < 2 < 3 < 4



Background and Objectives

Plots of average time trends (±se) per pixel

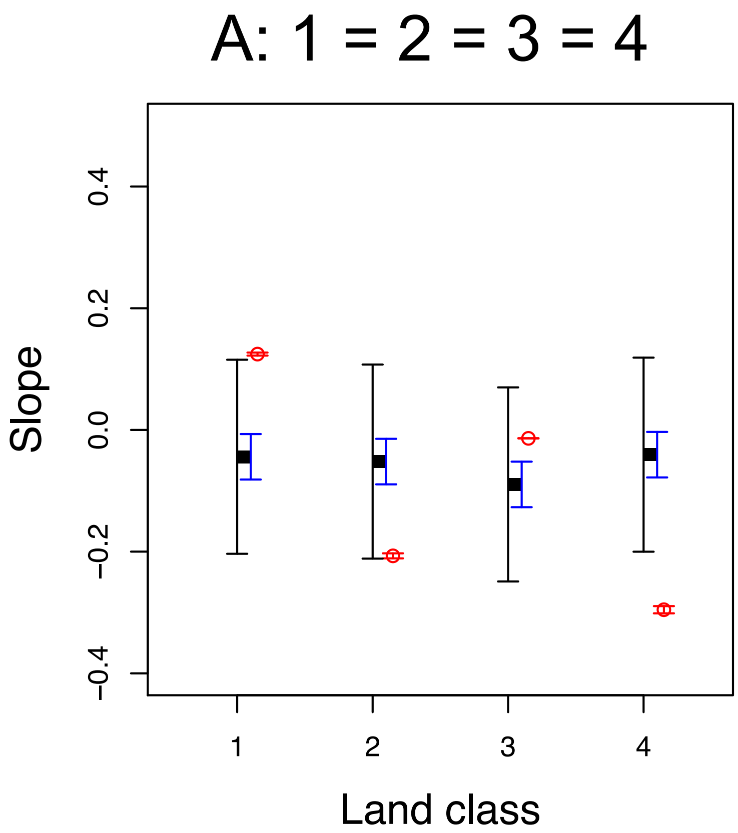


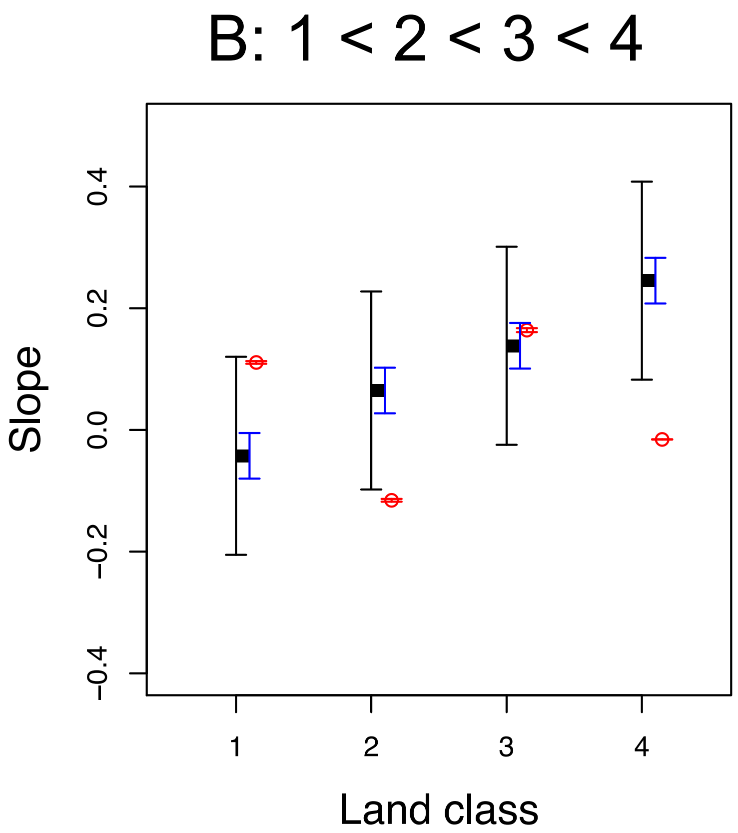


Background and Objectives

PARTS: Difference from zero

PARTS: Differences from each other





Background and Objectives

Example illustrations with simulations

1. Time trends vs. temporal autocorrelation
True time trends must be statistically distinguished from temporal autocorrelation

2. Spatial drivers vs. spatial autocorrelation
Spatial patterns do not always imply spatial drivers

3. Revealing true drivers underlying observed patterns
Accounting for spatial autocorrelation can reveal the true underlying patterns and allow them to be tested statistically

Background and Objectives

Objective

To develop statistically sound and computationally feasible ways to test hypotheses using remote-sensing data

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Technical and Science Advancements

Methods

1. Partitioned Autoregressive Time Series (PARTS)
2. Panel Regression (PR)

Applications

1. Global trends in NDVI (PARTS)
2. Changes in vegetation in Central Asian grasslands (PARTS)
3. Trends in Arctic summer sea ice (PARTS)

Two statistical approaches

1. PARTS ("old-school" statistics)

   i. Performs time-series analyses on individual pixels to estimate trends
   ii. Analyzes the spatial pattern in trends in random partitions of the map
   iii. Combines the tests from the partitions to give an omnibus statistical test for the map
   iv. Scales linearly to arbitrarily large maps
   v. Validated with simulations
   vi. R package remotePARTS

Technical and Science Advancements

Two statistical approaches

2. Panel regression ("new-school" statistics)

i. Uses a model of the complete system with time and space
ii. Creates a mathematically parsimonious representation of the full model
iii. Solves the complete time-and-space problem with quasi-maximum likelihood
iv. Scales linearly to arbitrarily large maps if spatial autocorrelation is moderate
v. Strong theoretical foundation for good statistical properties
vi. Validated with simulations
vii. Current implementation in MATLAB

Technical and Science Advancements

Two statistical approaches

Why two?

PARTS is “flexible” because it makes minimal assumptions about the data

PR is “stiff” because it relies on making more assumptions about a full time-and-space model

Statistical trade-off between robustness and power: a specific problem might be best-served by one or the other

By comparing both, we can improve both

Technical and Science Advancements

Methods

1. Partitioned Autoregressive Time Series (PARTS)
2. Panel Regression (PR)

Applications

1. Global trends in NDVI (PARTS)
2. Changes in vegetation in Central Asian grasslands (PARTS)
3. Trends in Arctic summer sea ice (PARTS)

Technical and Science Advancements

Methods

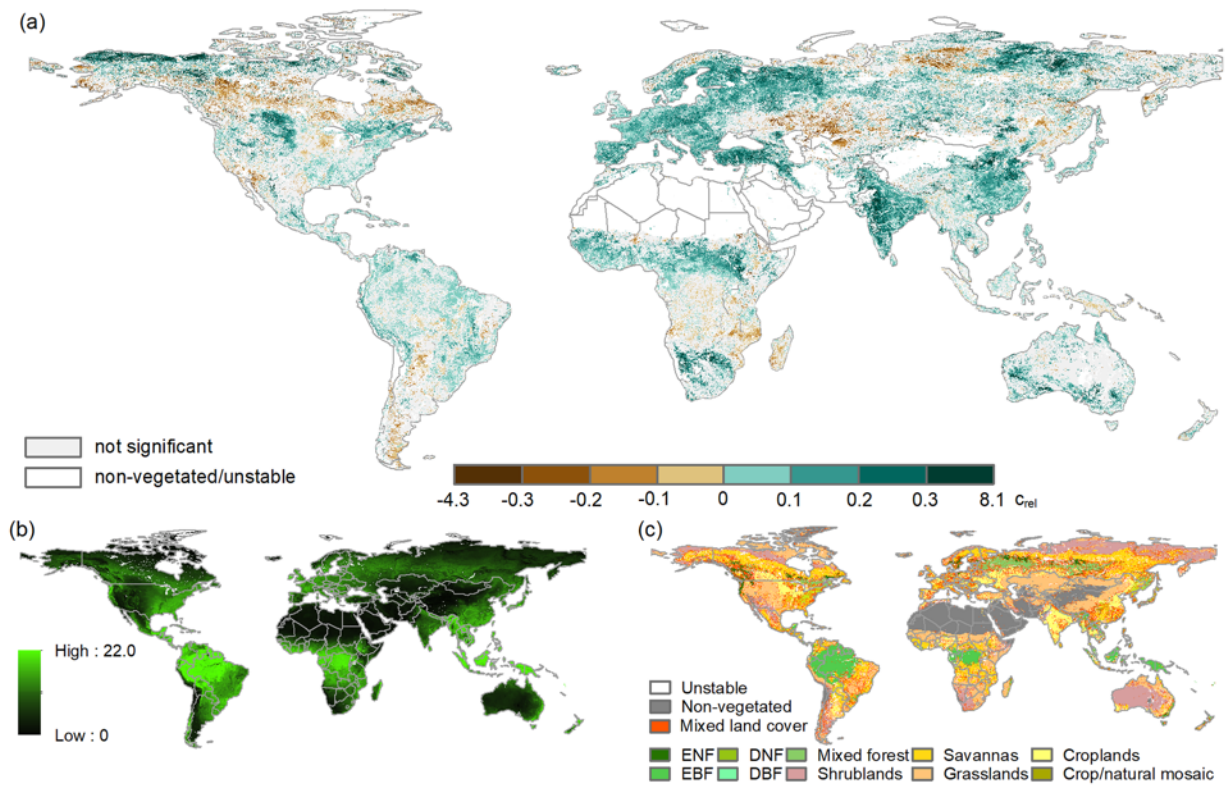
1. Partitioned Autoregressive Time Series (PARTS)
2. Panel Regression (PR)

Applications

1. Global trends in NDVI (PARTS)
2. Changes in vegetation in Central Asian grasslands (PARTS)
3. Trends in Arctic summer sea ice (PARTS)

Technical and Science Advancements

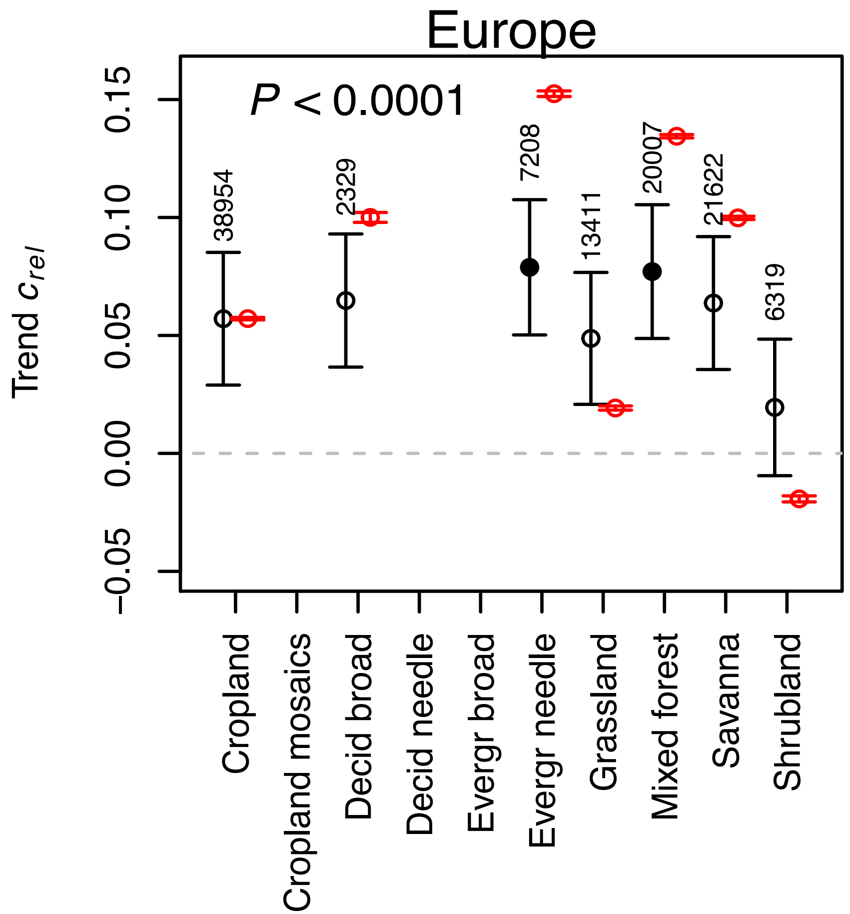
Global trends in NDVI 1982-2015 (PARTS, Likai Zhu)



Annual cumulative NDVI 1982-2015 from NDVI3g

MODIS land-cover product (MCD12Q1)

Differences in greening among land-cover classes



red = average trend (±se) per pixel

black = PARTS mean (±se)

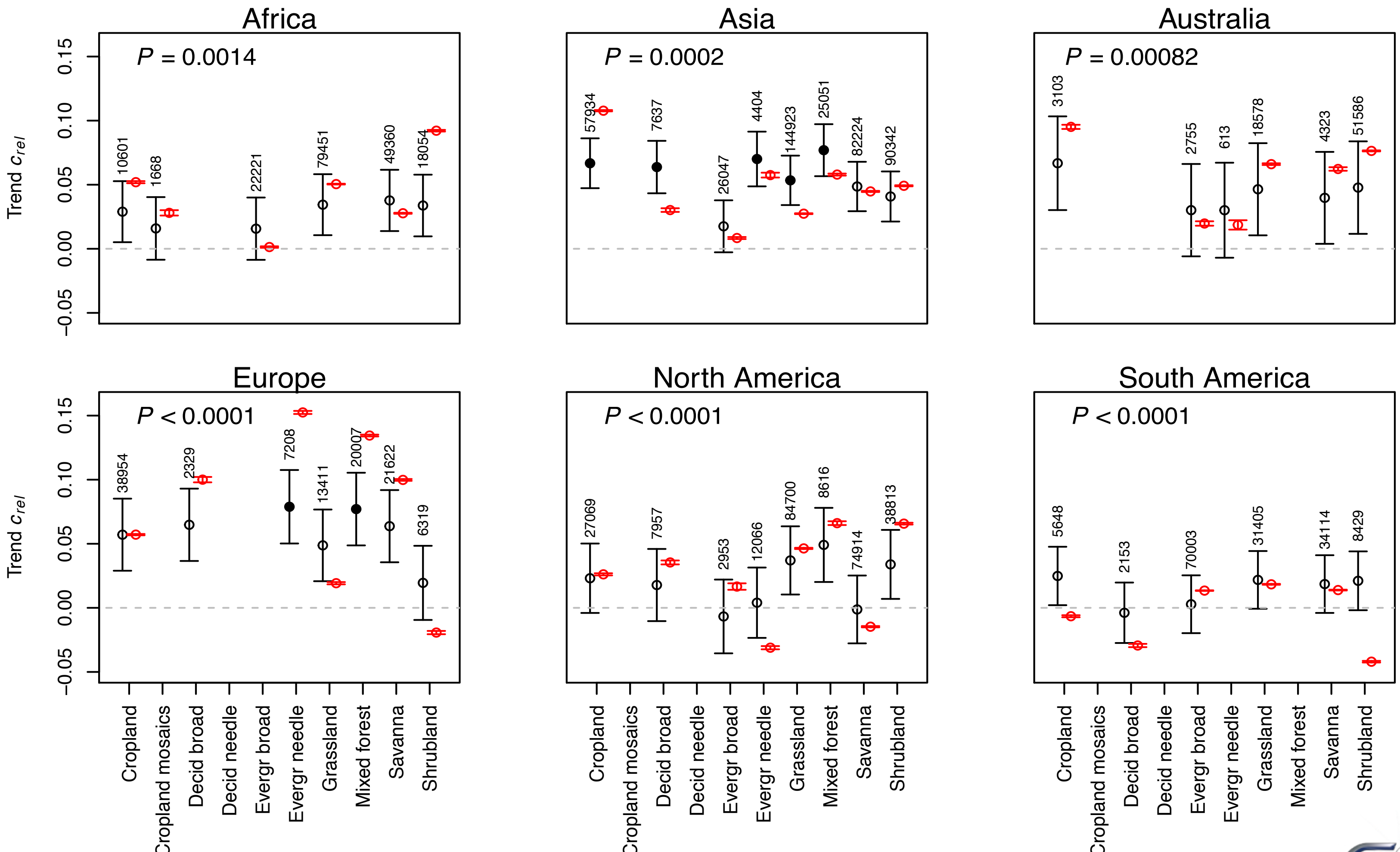
n = number of pixels

P for differences among land-cover classes

solid dot for differences from zero

Technical and Science Advancements

Differences in greening among land-cover classes



Differences in greening with latitude

Continent	n	P_{part}
Africa	214121	0.44
Asia	570250	1.00
Australia	91459	0.61
Europe	174498	0.08
North America	345264	0.75
South America	190832	0.15

Land-cover class by latitude interaction

Continent	n	P_{part}
Africa	182587	0.0074
Asia	441312	0.0004
Australia	80958	0.0555
Europe	110385	0.0006
North America	257344	0.0007
South America	152549	<0.0001

Technical and Science Advancements

Methods

1. Partitioned Autoregressive Time Series (PARTS)
2. Panel Regression (PR)

Applications

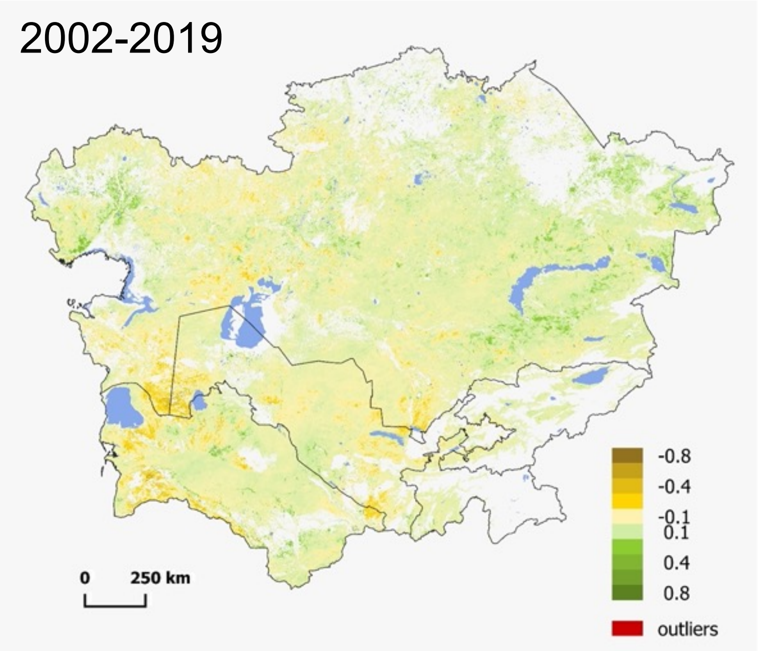
1. Global trends in NDVI (PARTS)
2. Changes in vegetation in Central Asian grasslands (PARTS)
3. Trends in Arctic summer sea ice (PARTS)

Technical and Science Advancements

Kasia Lewinska

Green vegetation Cumulative Endmember Fraction

(proxy for annual vegetation productivity)



Range of spatial autocorrelation: 125 km

Land-cover classes differ
($F_{2,123954}$ = 2.369, P = 0.00001)

Weak differences among ecoregions
($F_{5,118270}$ = 1.44, P = 0.04)

ecoregions: Olson et al. (BioScience, 2001)

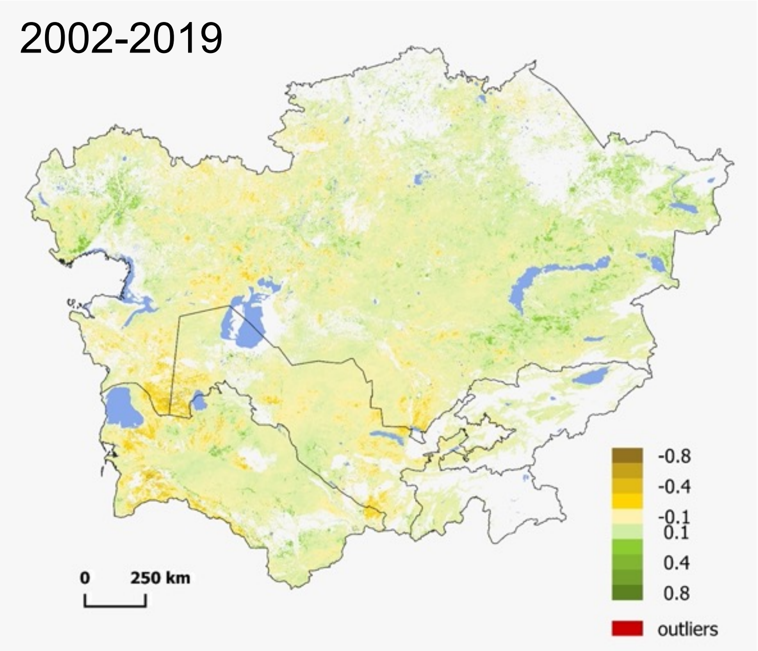
AGU 2020 Fall meeting (10.1002/essoar.10505216.1)

Technical and Science Advancements

Kasia Lewinska

Rain-Use Efficiency

(indicates ecosystem functioning)



Range of spatial autocorrelation: 305 km

Land-cover classes differ
($F_{2,124062} = 6.615$, $P = 0.00001$)

No differences among ecoregions
($F_{5,118270} = 0.72$, $P = 0.95$)

AGU 2020 Fall meeting (10.1002/essoar.10505216.1)

Technical and Science Advancements

Methods

1. Partitioned Autoregressive Time Series (PARTS)
2. Panel Regression (PR)

Applications

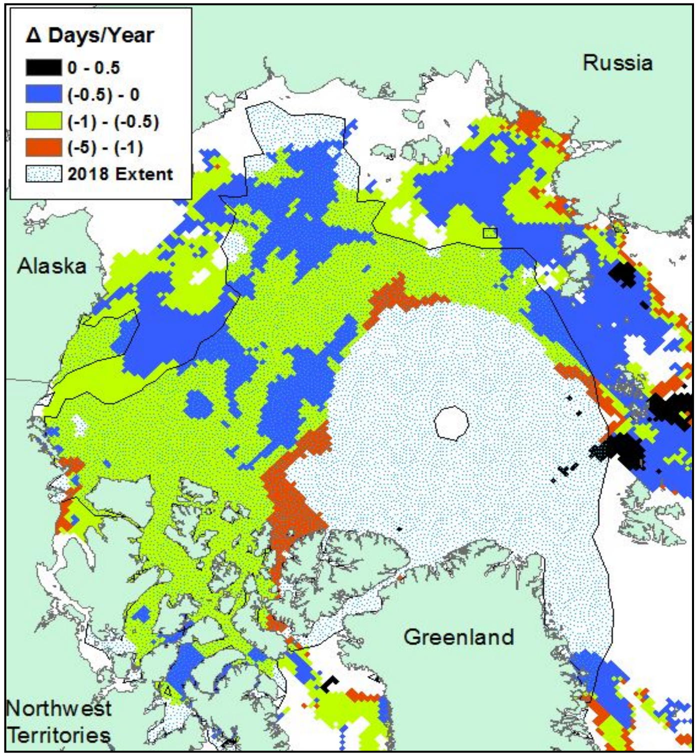
1. Global trends in NDVI (PARTS)
2. Changes in vegetation in Central Asian grasslands (PARTS)
3. Trends in Arctic summer sea ice (PARTS)

Technical and Science Advancements

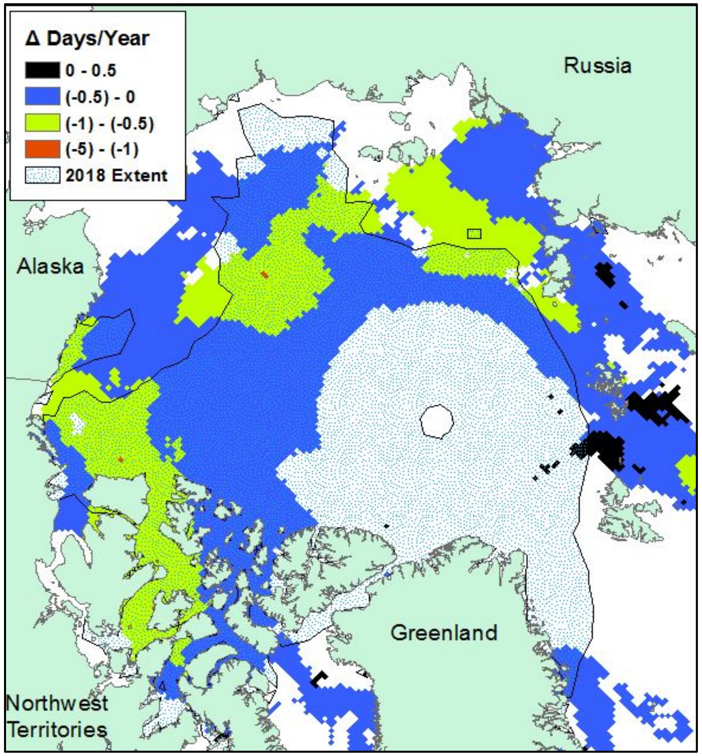
Connor Stephens

Loss of days of ice cover in September

No temporal autocorrelation
(least-squares regression)



Regression with temporal
autocorrelation



Technical and Science Advancements

Methods

1. Partitioned Autoregressive Time Series (PARTS)
2. Panel Regression (PR)

Applications

1. Global trends in NDVI (PARTS)
2. Changes in vegetation in Central Asian grasslands (PARTS)
3. Trends in Arctic summer sea ice (PARTS)

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Summary of Accomplishments and Future Plans

Current state

1. remotePARTS (R package: Clay Morrow)
   As we are applying it to different datasets and questions, we are learning and modifying remotePARTS accordingly.

2. Panel regression (Ralph Ting Fung Ma)
   The theoretical development is largely done, and the next step is challenging it with datasets.

   [COVID-19 lag in PR development]

Summary of Accomplishments and Future Plans

Future plans (methods)

Comparing PARTS and panel regression

Spatiotemporal drivers (independent variables change through time within a pixel)

For example, estimating the effect of changing temperature (independent variable) on changing green-up phenology

Future plans (applications)

Analyses	Satellite dataset	Who is in charge	Current state
Global greening and browning since 1981	GIMMS NDVI3g	Tony (PARTS)	Analyses complete, manuscript written
Overgrazing in arid regions	MODIS/in-house	Kasia (PARTS)	Compiling dataset for Central Asia
Change in sea-ice	NOAA	Connor (PARTS)	Datasets complete, pixel-level trends started
Surface temperature and urban heat islands	MODIS	Ralph (PR)	Datasets complete
Change in phenology	MODIS Veg. dynamics	Clay (PARTS and PR)	Datasets complete
Boreal forest fires	Landsat	Connor (PARTS)	Datasets being assembled
Trends in Dynamic Habitat Indices	MODIS/in-house	Lena (PARTS)	Datasets complete

Presentation Contents

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

Publications

Journal / Conference Papers:

Ives, A. R., and C. Bozzuto. 2021. Estimating and explaining the spread of COVID-19 at the county level in the USA. Communications Biology **4**:60.

Ives, A. R., L. Zhu, F. Wang, J. Zhu, C. J. Morrow, and V. C. Radeloff. in review. Statistical inference for spatiotemporal trends in remote-sensing data. Remote Sensing of Environment.

Ives, A. R. in review. Statistical tests for non-independent partitions of large autocorrelated datasets. MethodsX.

Presentations

Lewinska, K. E., A. R. Ives, C. J. Morrow, N. Rogova, and V. C. Radeloff. 2020. Monitoring trends in grasslands in Central Asia while accounting for temporal and spatial autocorrelation. American Geophysical Union.

Ives, A. R., L. Zhu, F. Wang, J. Zhu, C. J. Morrow, and V. C. Radeloff. 2020. Statistical inference for spatiotemporal trends in remote-sensing data. American Geophysical Union.

Acronyms

List of Acronyms

- PARTS — Partitioned Autoregressive Time Series model
- PR — Panel regression

An Analytic Center for Biodiversity and Remote Sensing Data Integration

Walter Jetz (PI, Yale University)
Adam Wilson (Co-I, University at Buffalo)
Rob Guralnick (Co-I, University of Florida)

AIST-18-0034 Annual Technical Review
5 February 2021

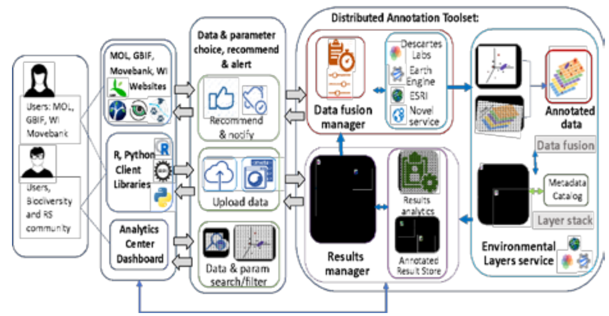
An Analytic Center for biodiversity and remote sensing data integration

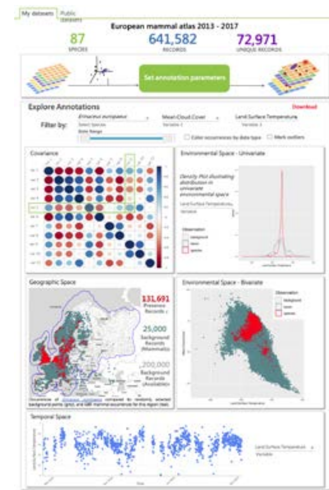
PI: Walter Jetz, Yale University

Objective

- Increase the usability and accessibility of linked remote sensing (RS) and biodiversity information in support of both communities through a versatile analysis and support dashboard, client libraries and shared products
- Provide pre-computed annotations of >1B biodiversity observations for a large suite of RS or RS-derived products, together with dynamic updating.
- Support the informed integration of biodiversity and RS data by providing data-driven guidance regarding respective spatiotemporal grains and most effective combination of RS information for downstream use.
- Improve documentation of environmental biodiversity data coverage and its signal about biodiversity responses to rapid environmental change.

Workflow schematic





Online dashboard

Approach:

- Leverage prototyped software tools to connect Map of Life and multiple remote sensing layer sources (Google Earth Engine, Descartes Labs, ESRI).
- Apply fusion tools to existing and user-provided spatial biodiversity data and a set of remote sensing layers.
- Create an online dashboard and client libraries for user-driven custom annotation with visual reporting tools to support variable selection, background characterization, and trends in environmental niches.
- Contribute environmental metrics, reports, and graphics to partners (MOL, GBIF, Movebank, Wildlife Insights).

CoIs: Adam Wilson (U Buffalo)
Rob Guralnick (U Florida)

Key Milestones

- *Data connectors to the three platforms complete 09/20*
- *Pre-annotations (for data up to that time) executed 12/20*
- *Environmental data reports available on MOL 03/21*
- *Final R library and documentation published on CRAN 12/21*

TRL_{in} = 5 **$TRL_{current}$ = 6**

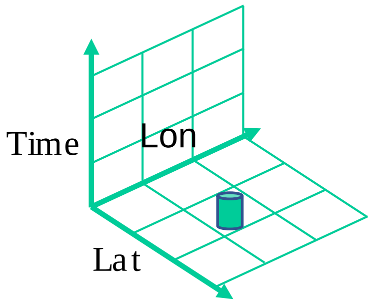
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background, Objectives

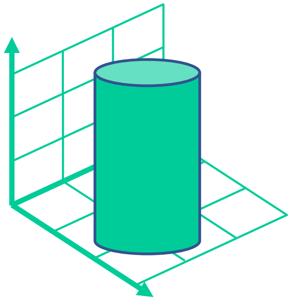
- For many uses in ecological forecasting and biodiversity science the **integration** of remote sensing and biodiversity data is **too complex and inflexible or lacks transparency and guidance**
- This Analytic Center addresses this need through a combination of user-friendly **dashboards** linked to a scalable environmental annotation **engine** and vast **precomputed information**
- It supports the informed integration of biodiversity and RS data by providing **data-driven guidance** for the selection of **variables** and **spatiotemporal grains** to characterize **environmental niches**
- The Center enables improved documentation of **environmental biodiversity data coverage** and its **signal** about biodiversity responses to **rapid environmental change**.

Background, Objectives

Grain of record

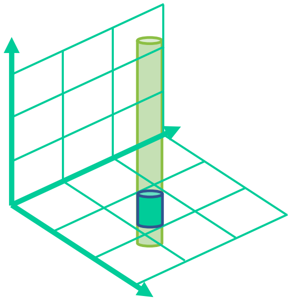


Camera trap record

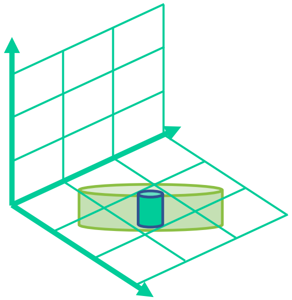


National park inventory

Extent of driver

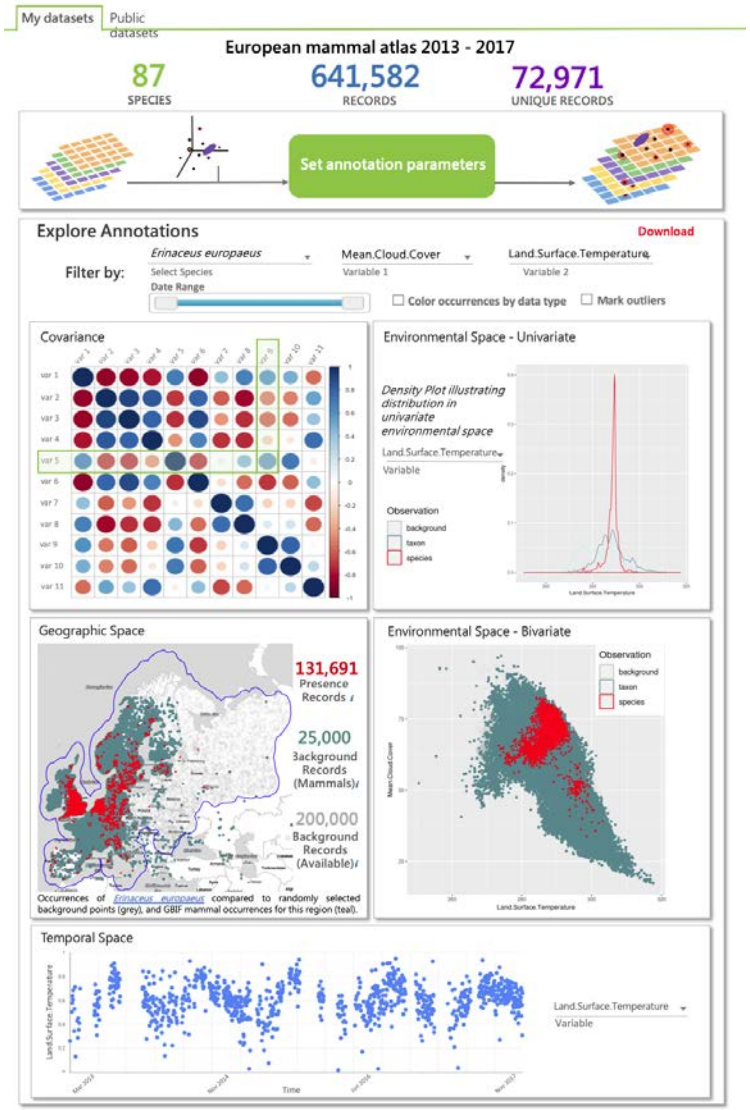


Flowering onset



Landscape drivers of presence

Background, Objectives



Dashboard from Proposal

User Perspective:

- Quantify the environmental niche of species, offer contrast to background or broader taxonomic group
- Increase *environmental data* options to include more remote sensing data
- Increase the options for spatial and temporal aggregation
- Support and provide guidance for data analysis and variable selection,

Background, Objectives

1. Leverage prototyped software tools to connect MOL and multiple RS layer sources (Google Earth Engine, Descartes Labs, ESRI).

2. Apply fusion tools to existing and user-provided spatial biodiversity data and a set of remote sensing layers.

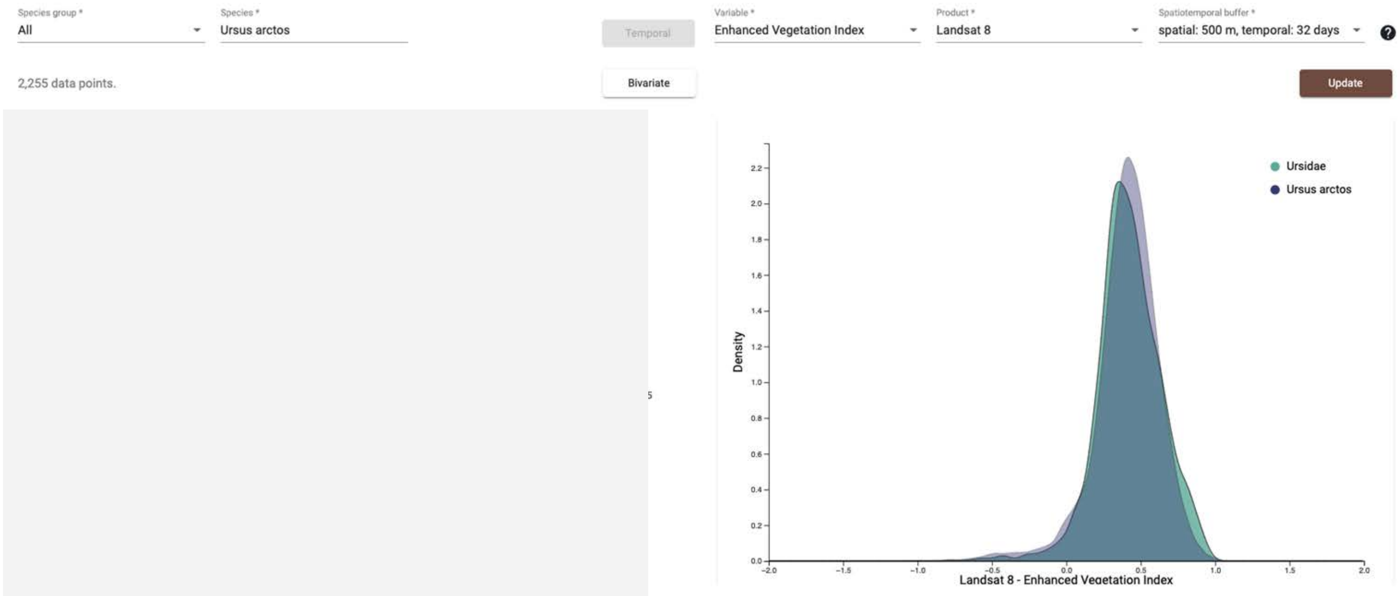
3. Create an online dashboard and client libraries for user-driven custom annotation.

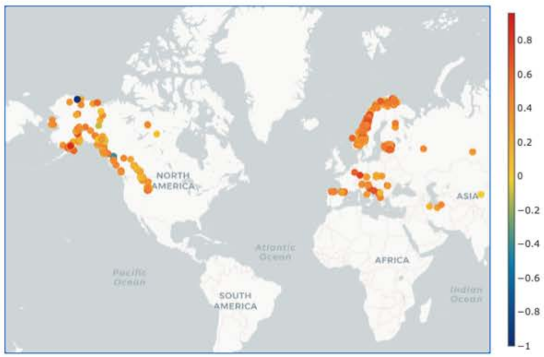
4. Contribute environmental metrics, reports, and graphics to partners (MOL, GBIF, Movebank, Wildlife Insights).

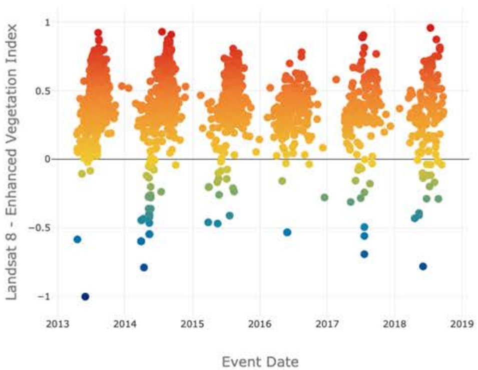
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Technical and Science Advancements

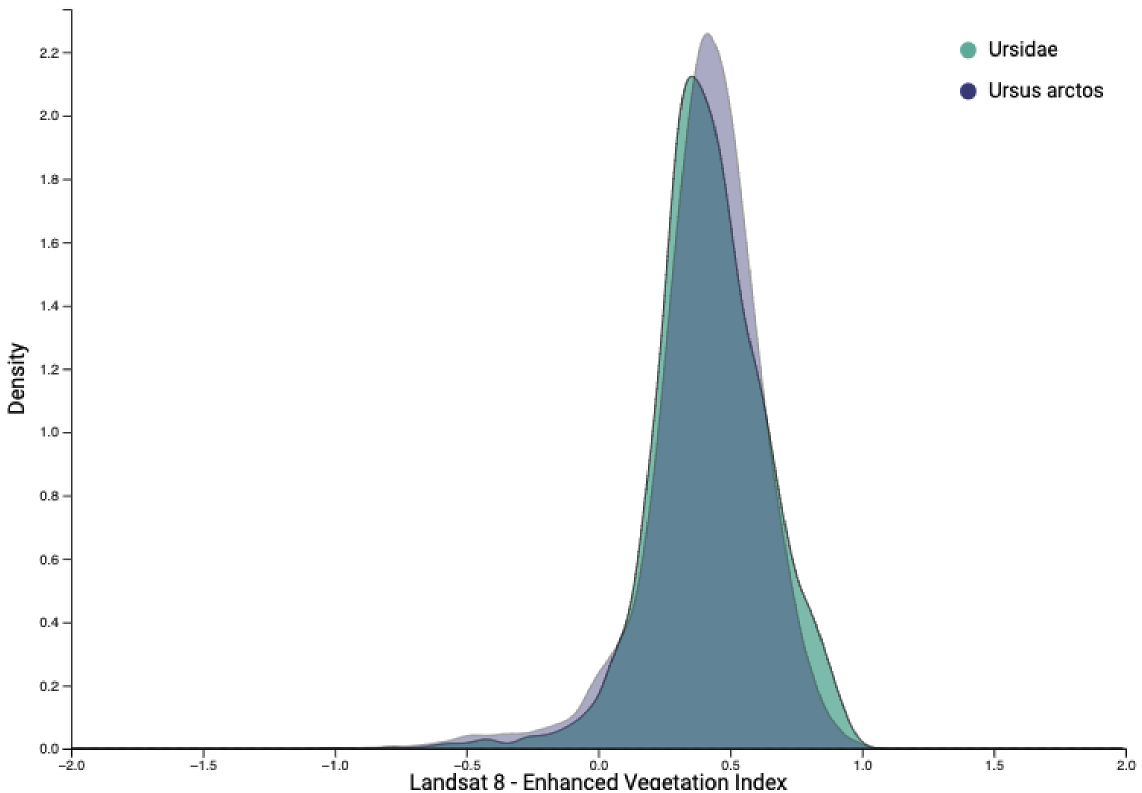
Dashboard Components: Current Status







Technical and Science Advancements



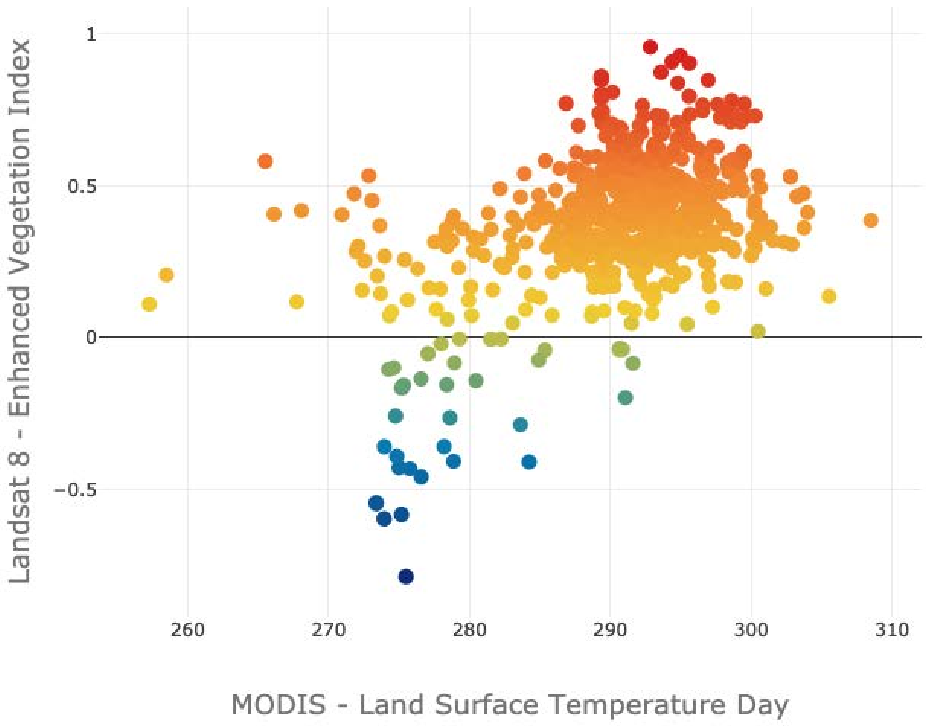
Display species environmental niche for single variable (density).

Optionally: contrast with broader taxonomic group

Genus, Family, Order

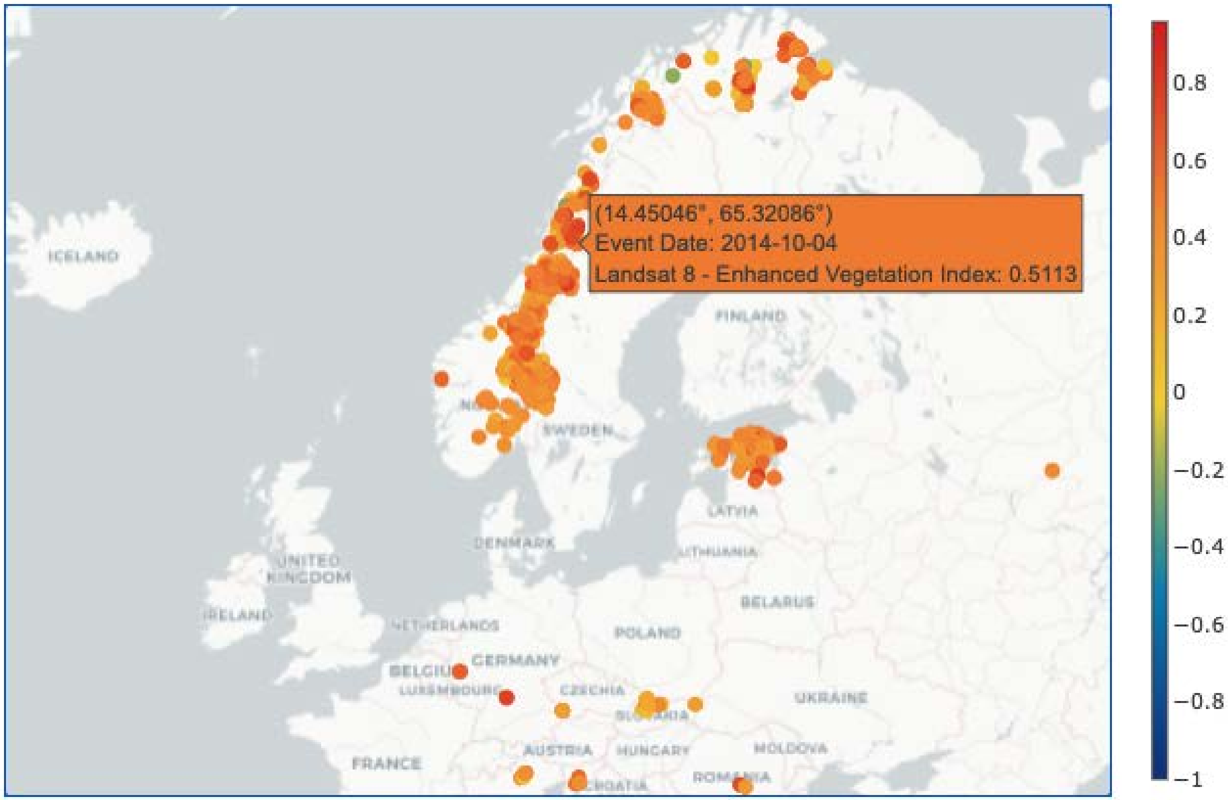
Helps user understand environmental differences compared to related species

Technical and Science Advancements



Display species environmental niche for two variables

Technical and Science Advancements

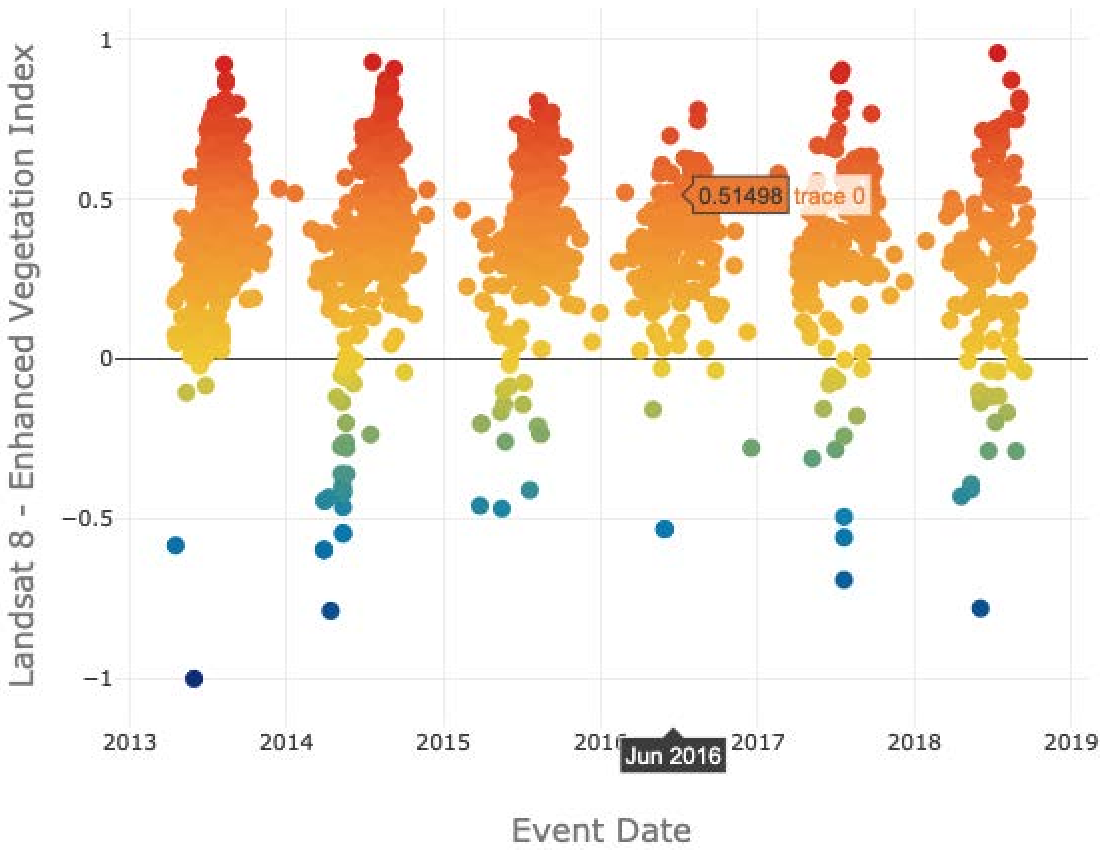


Species occurrences annotated with environmental conditions

Helps user understand spatial variability in environmental conditions for the species of interest

ESTO Earth Science Technology Office

Technical and Science Advancements

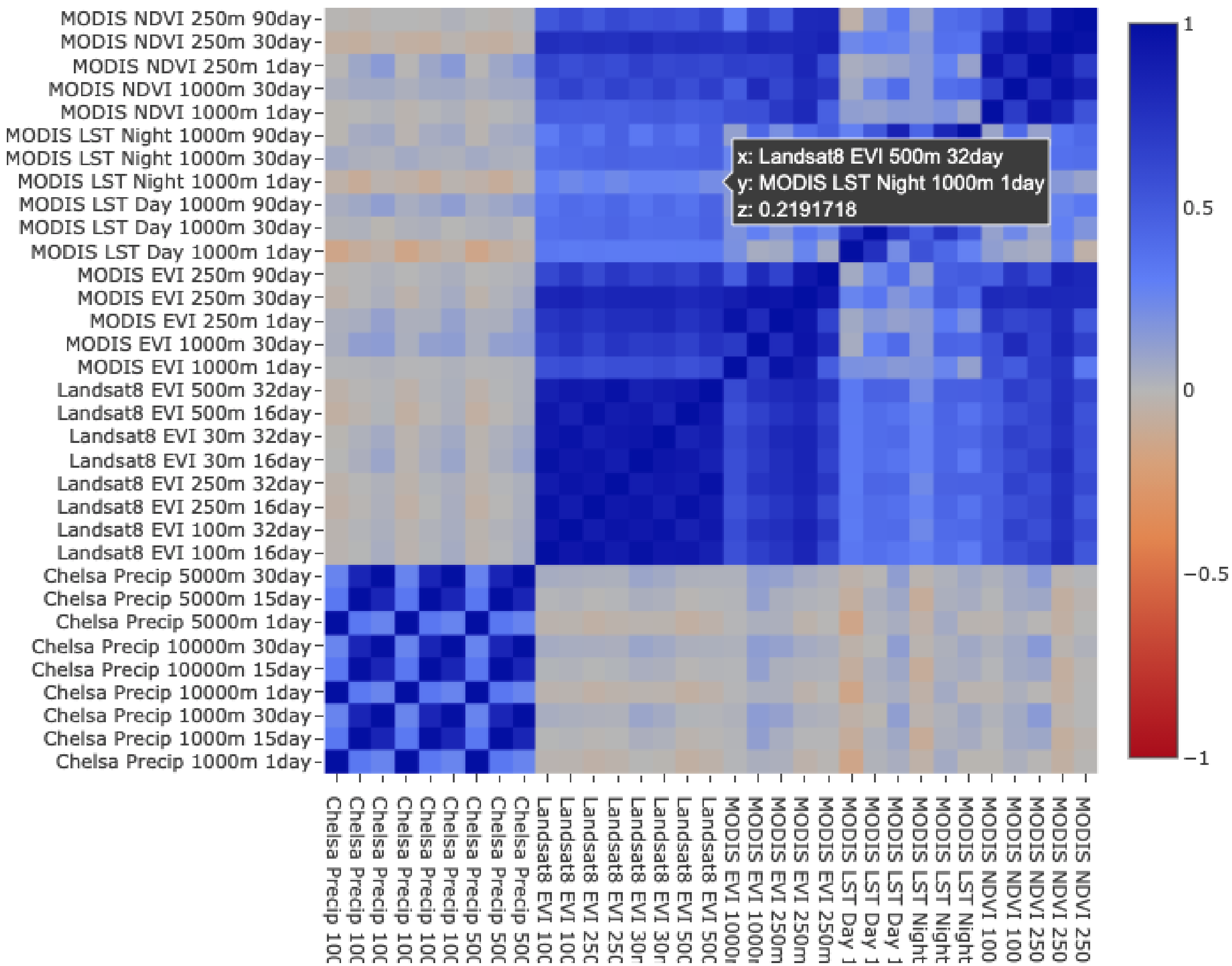


Temporal distribution of environmental conditions at occurrence locations

Helps user understand temporal variability (seasonality & interannual trends) in environmental conditions for the species of interest

Technical and Science Advancements

2,255 data points   Ursus arctos

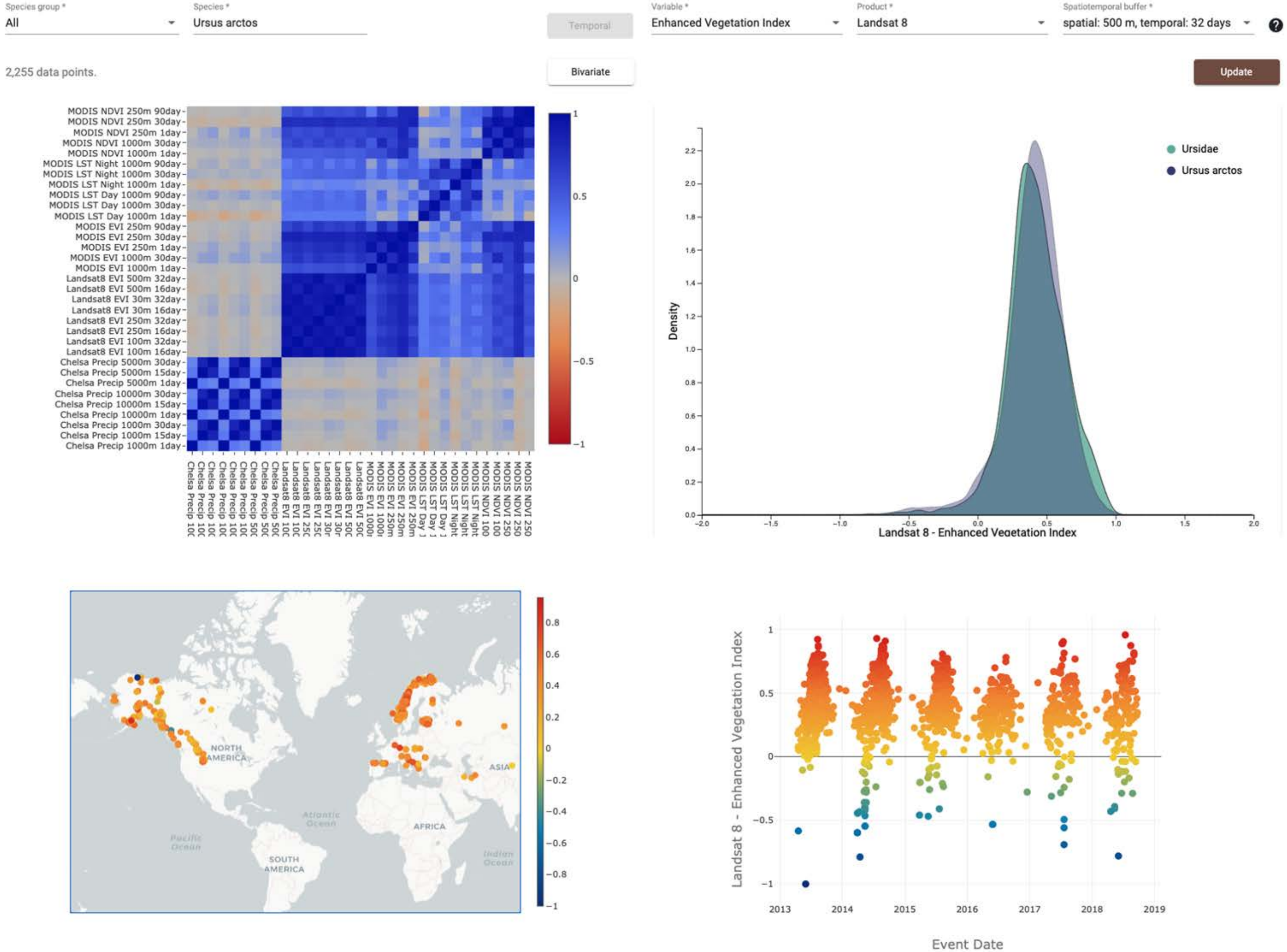


Exploratory Data Analysis and Variable Selection

Visualize relationships between environmental datasets to guide user through variable selection

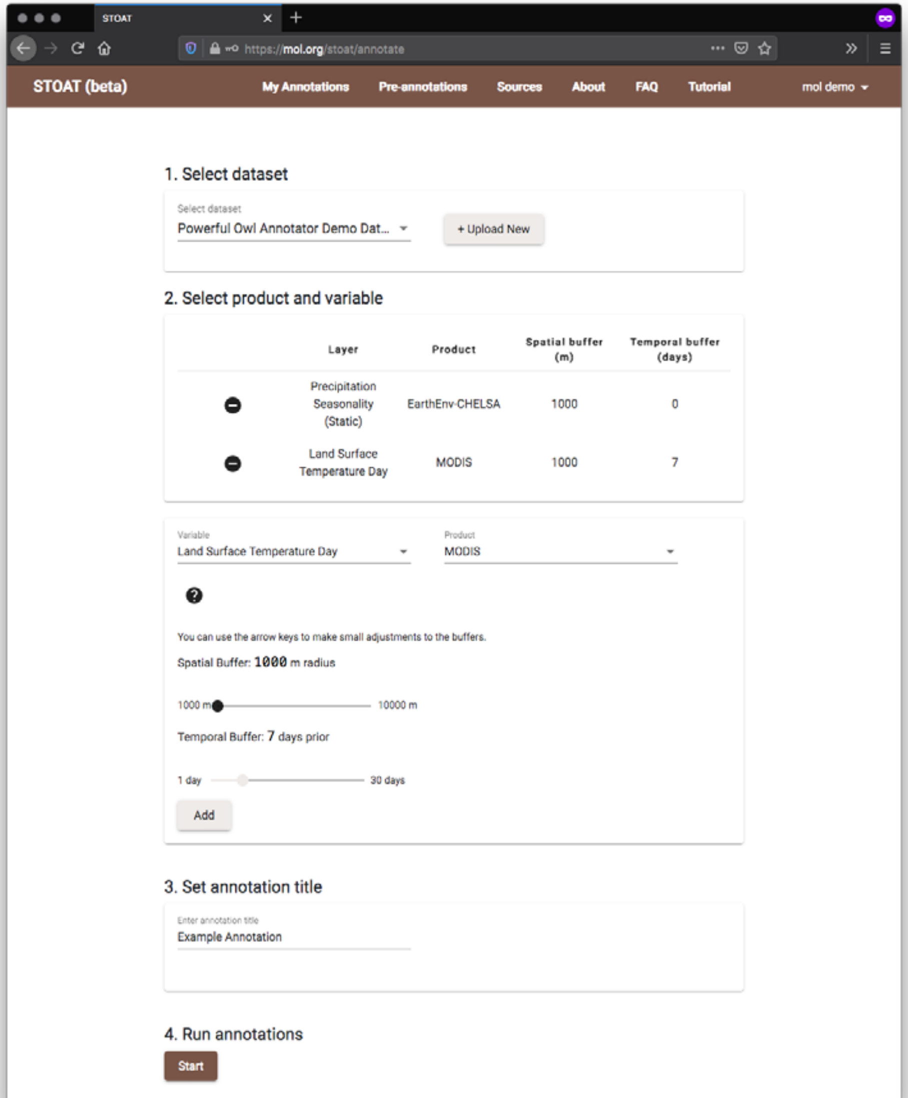
To explore: on selection of variable, change of colors to show correlations selected variable

Dashboard Components: Current Status



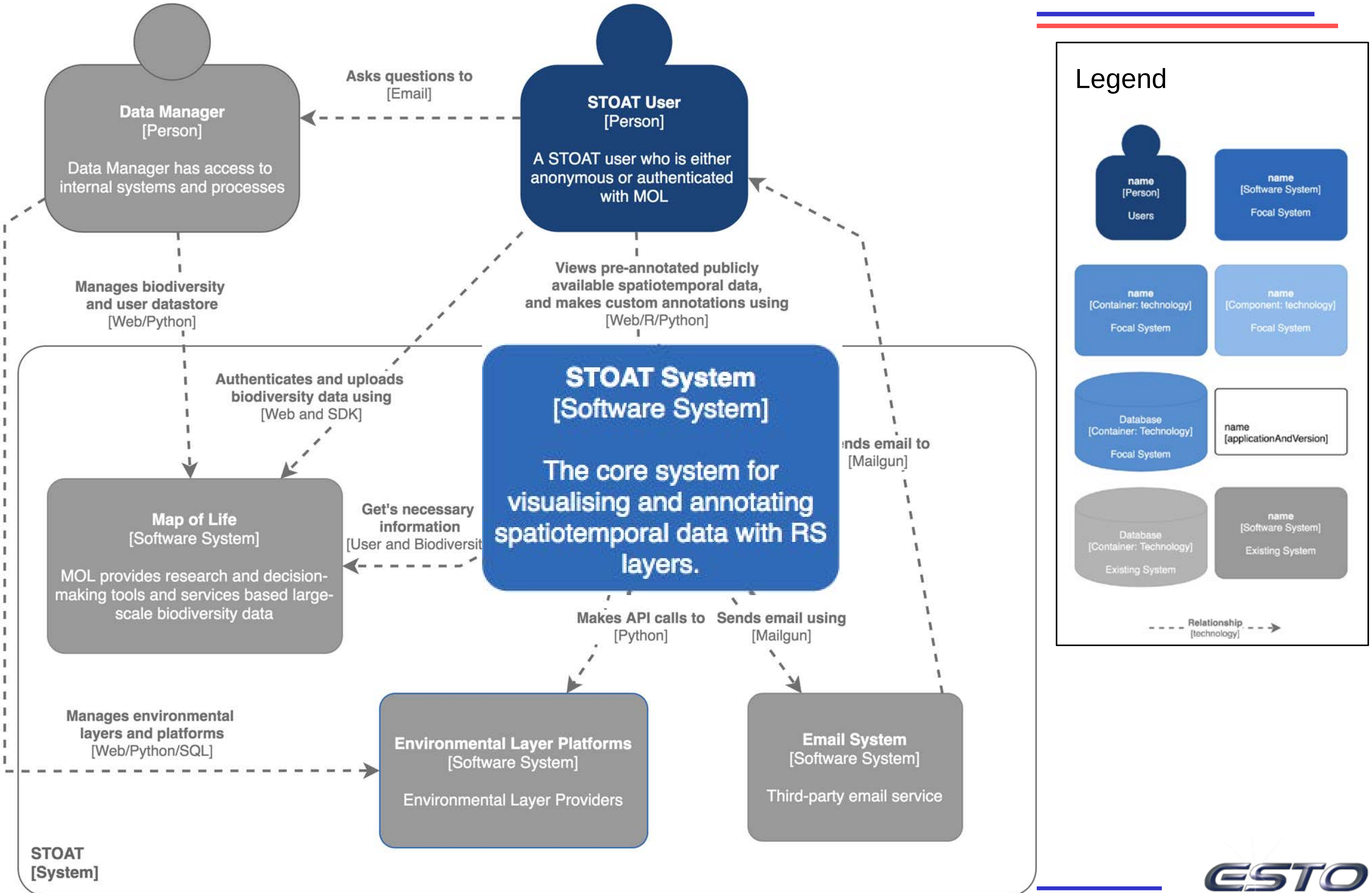
Earth Science Technology Office

Technical and Science Advancements

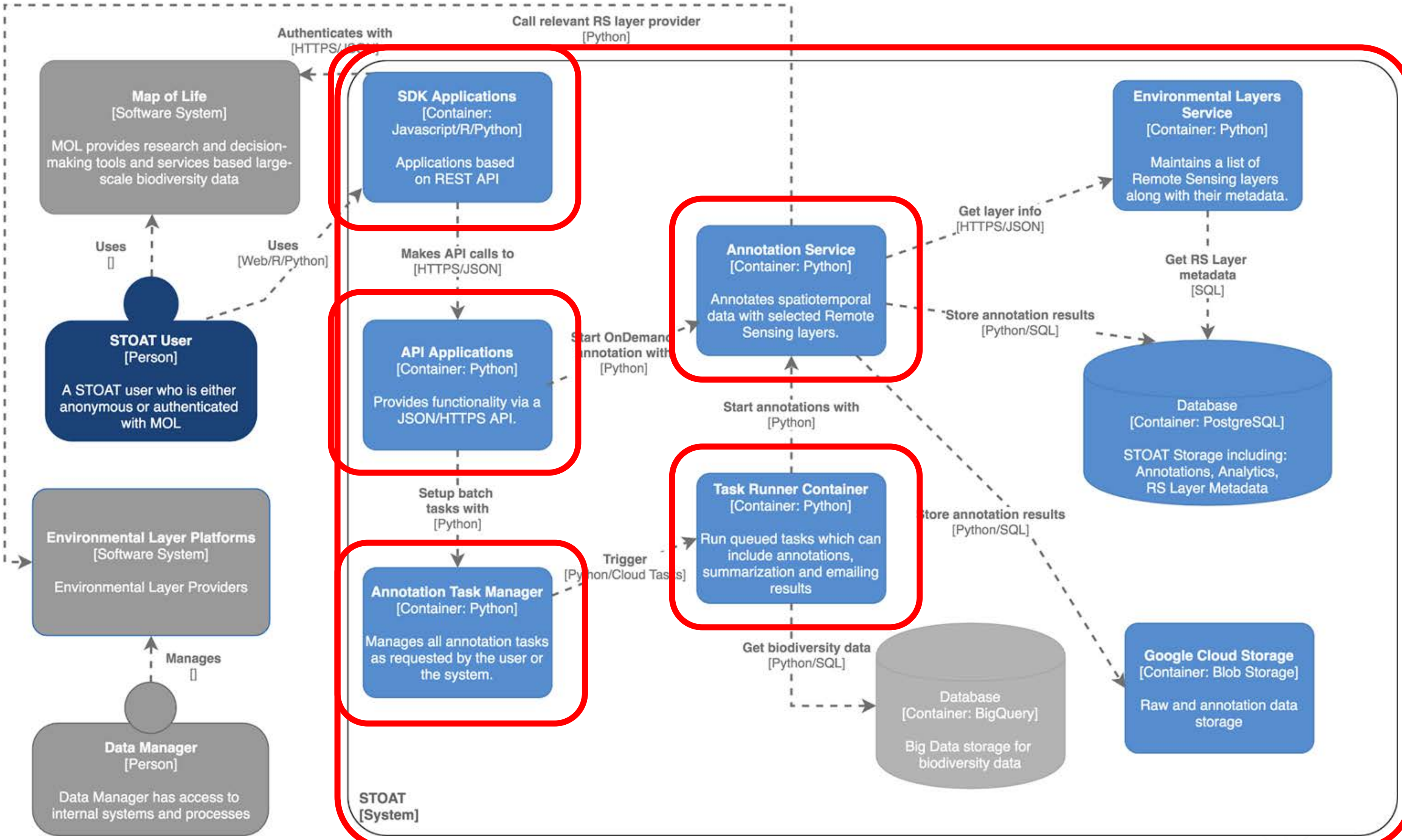


- Custom Annotation Dashboard
  - Completed UI
  - Customizable spatial and temporal buffers
  - Many static layers added to layer list
- Custom annotation for static layers in production
  - Includes improved custom annotation workflow

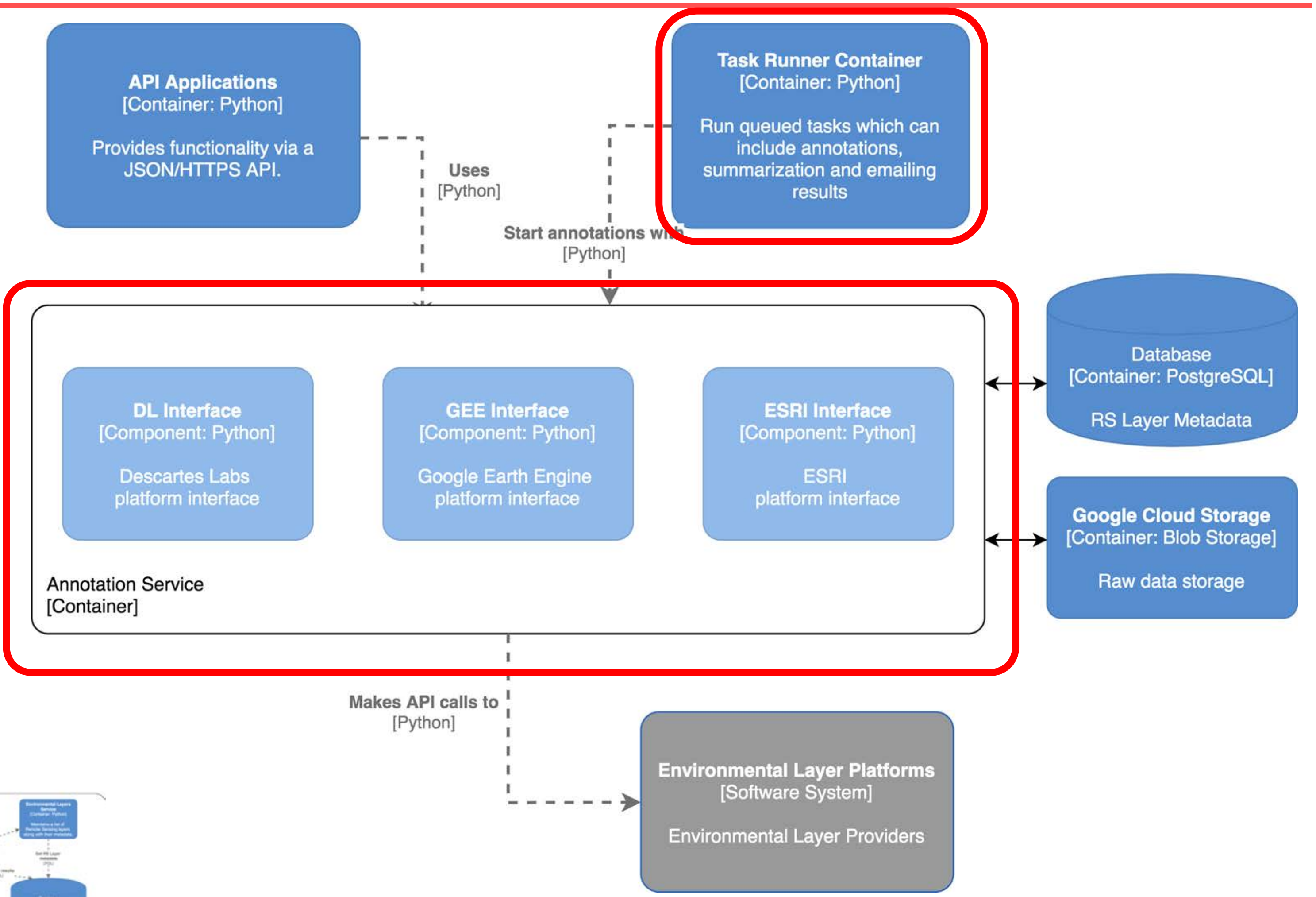
Technical and Science Advancements

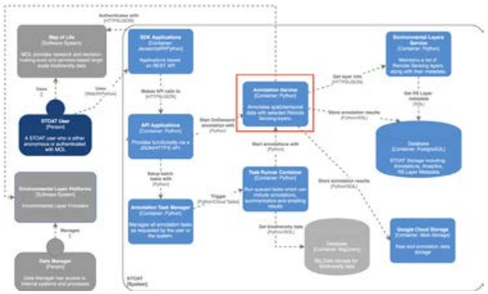


Technical and Science Advancements

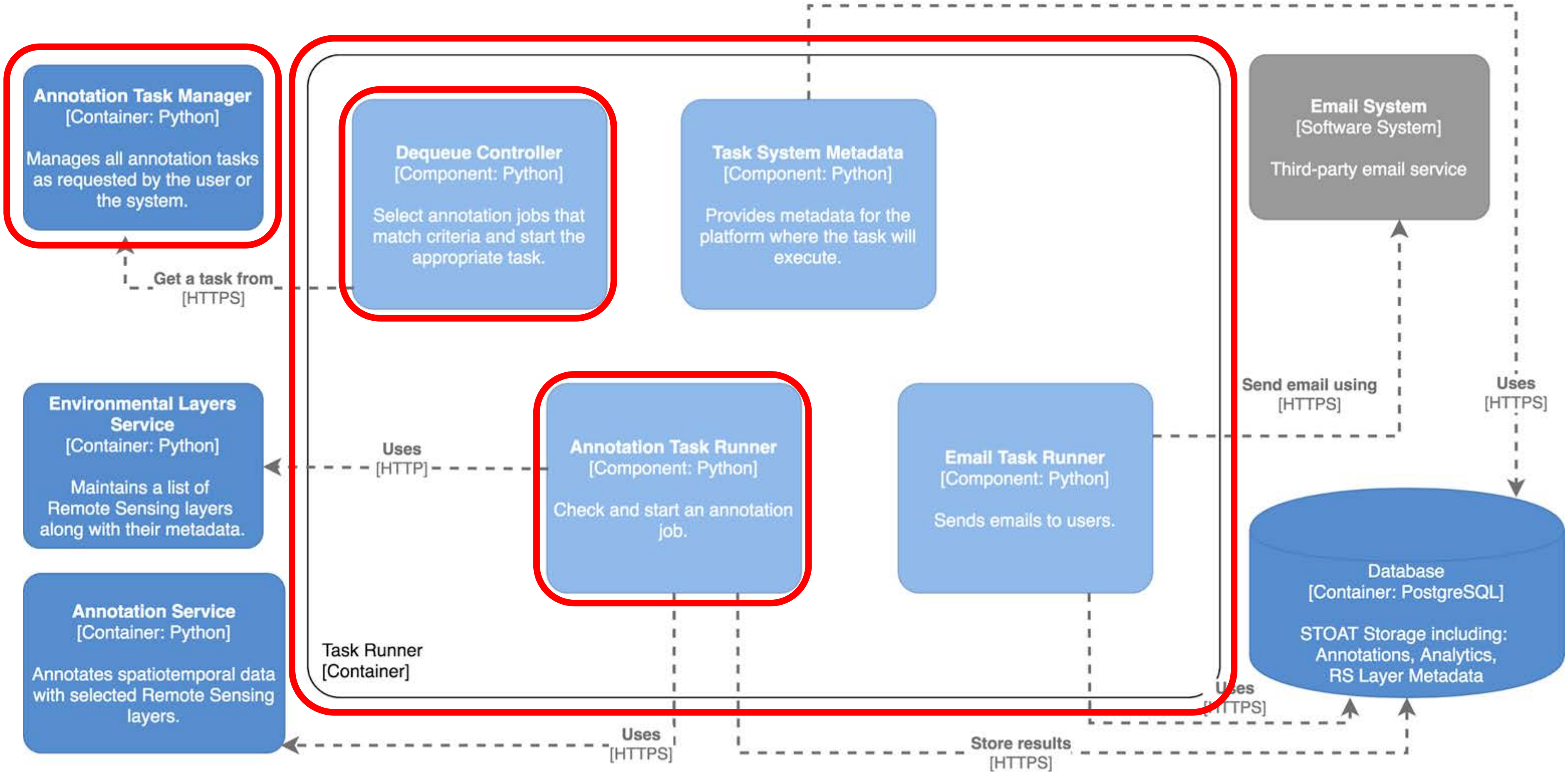


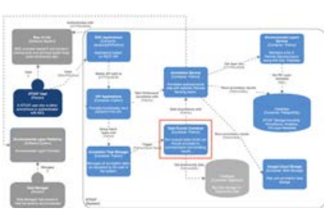
Technical and Science Advancements



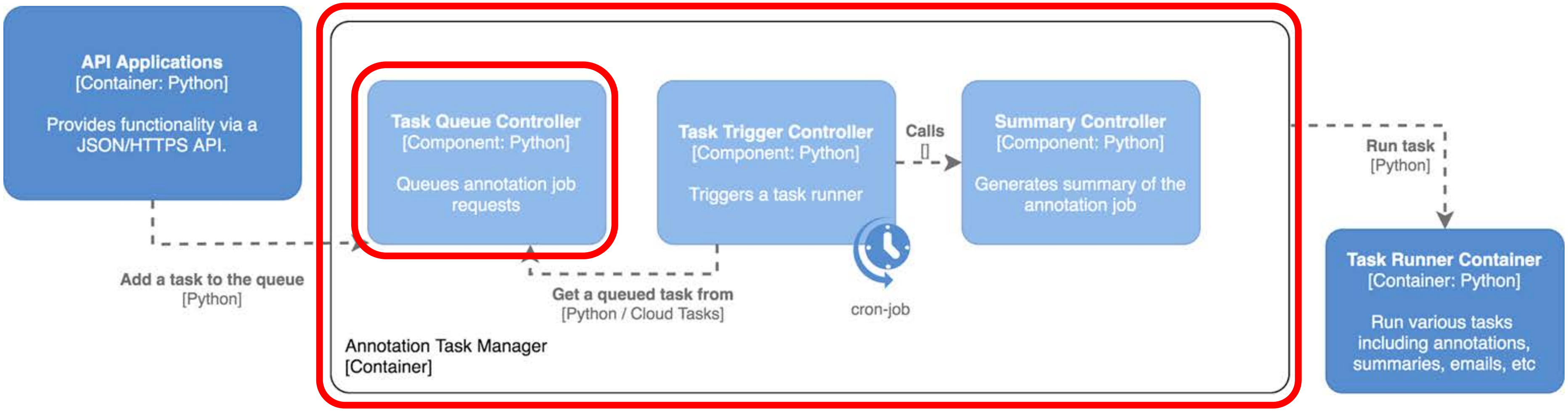


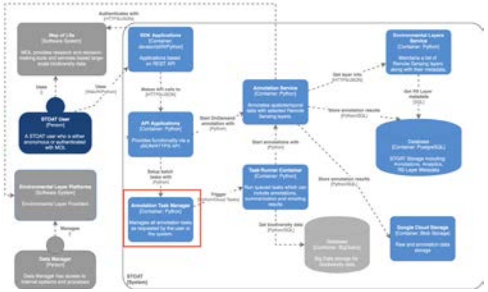
Technical and Science Advancements



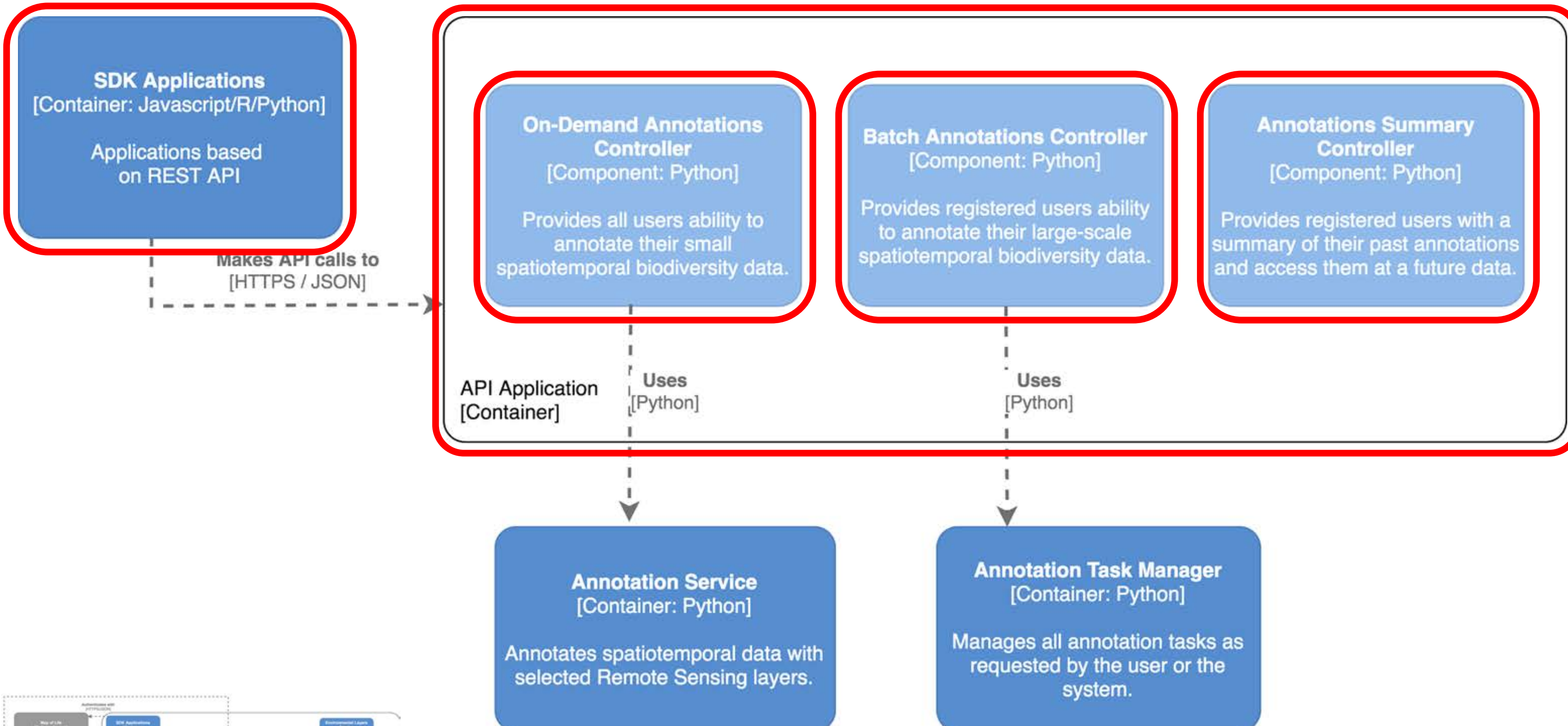


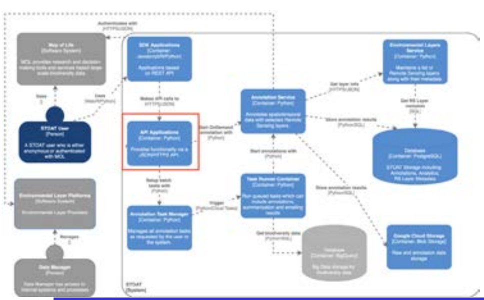
Technical and Science Advancements



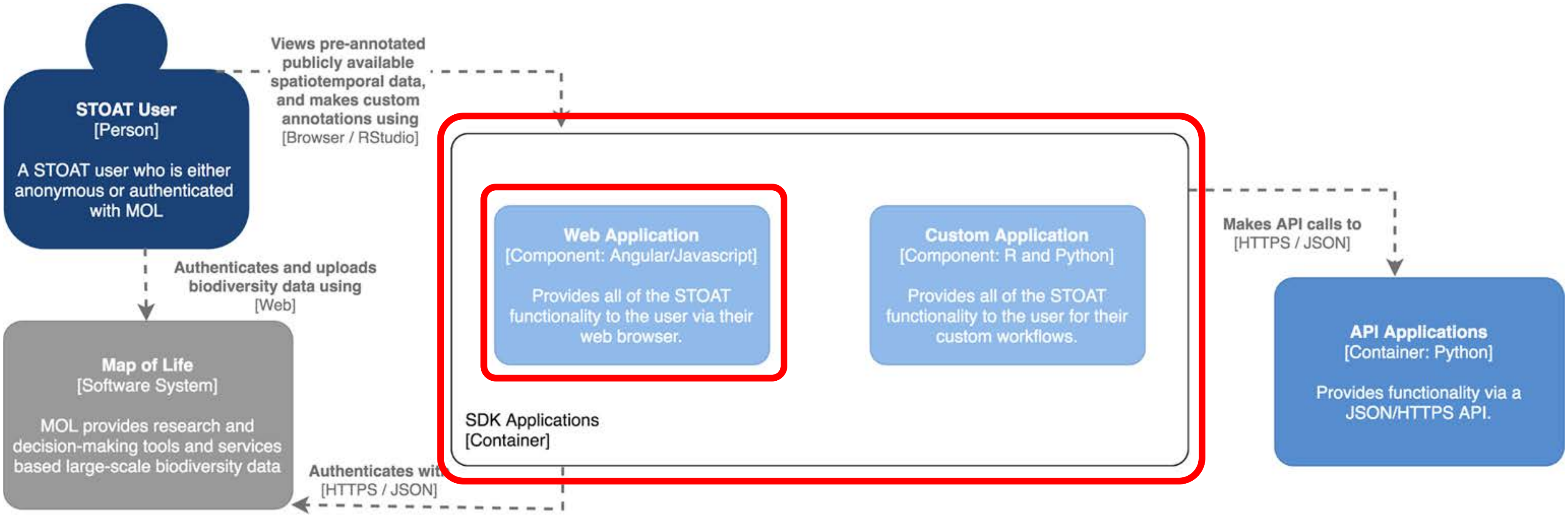


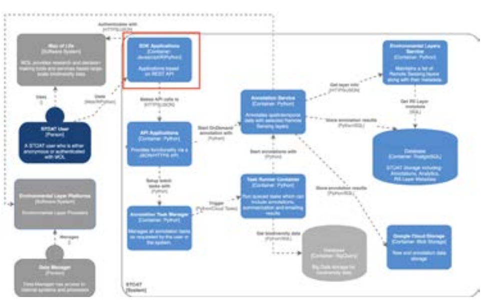
Technical and Science Advancements

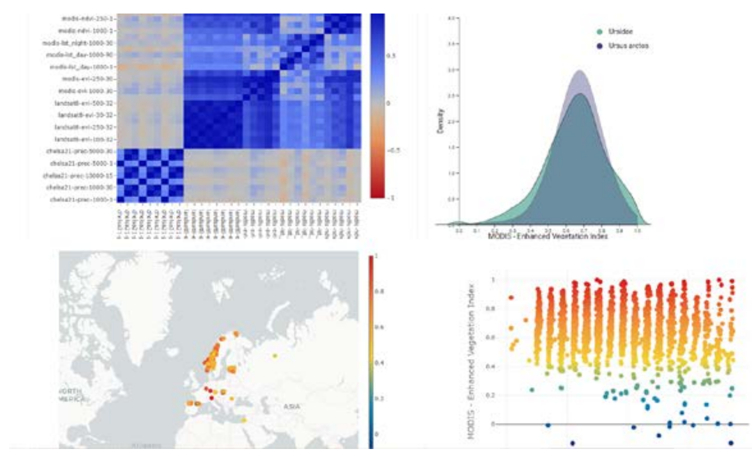




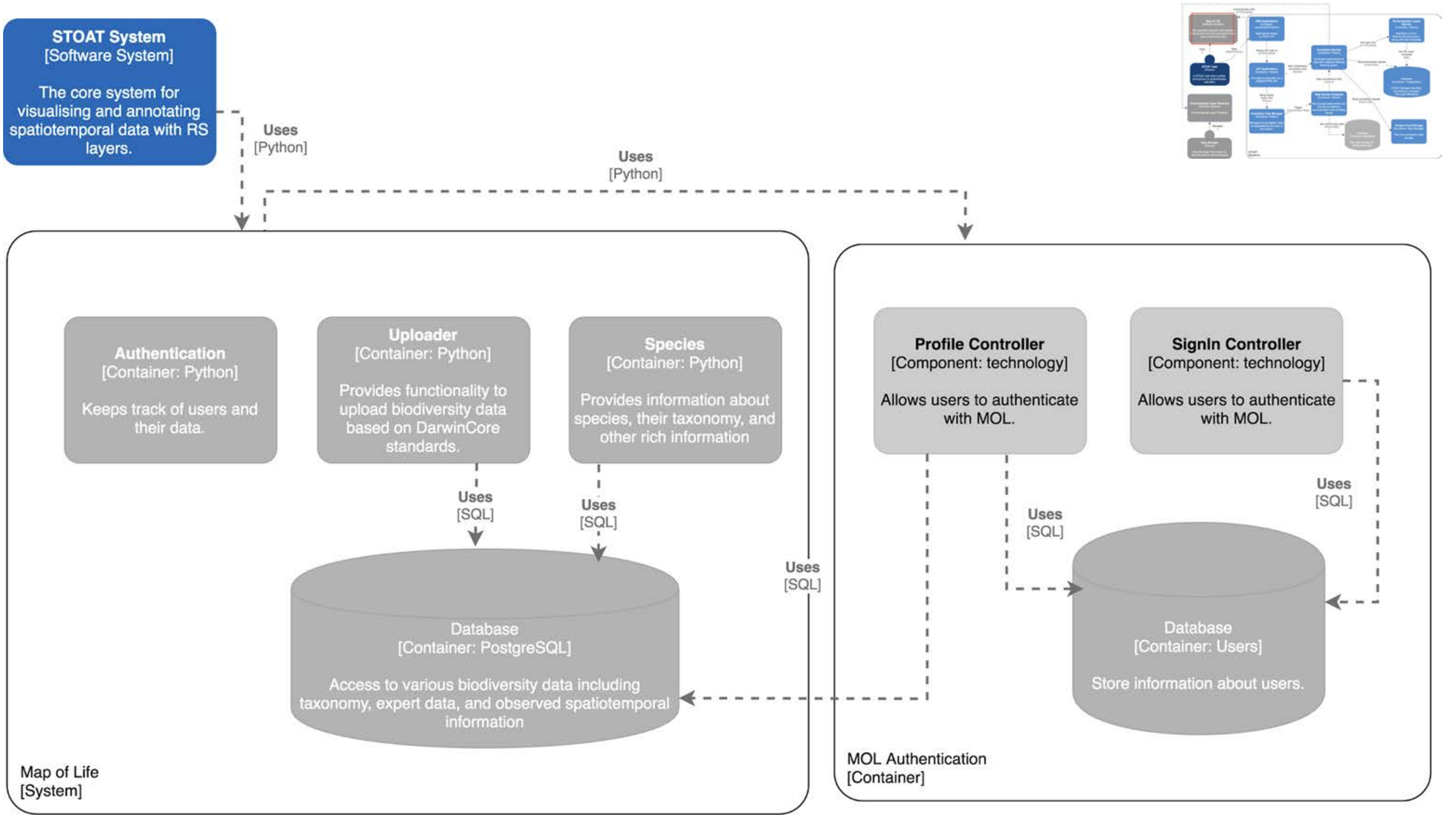
Technical and Science Advancements



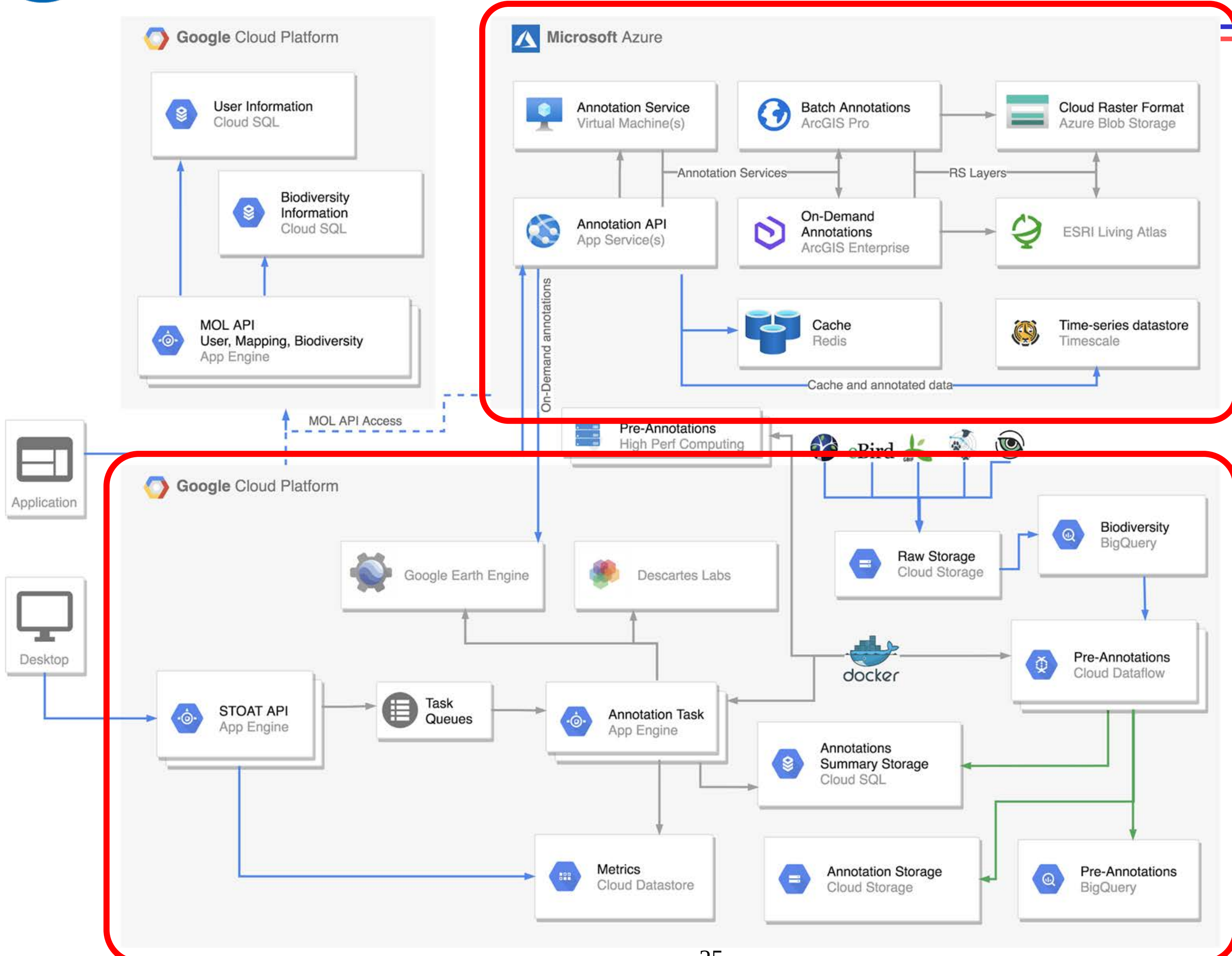




Technical and Science Advancements



Technical and Science Advancements



Technical and Science Advancements

Count of annotated species occurrence records

Annotation Run	Amphibians	Birds	Fishes	Invertebrates	Mammals	Other	Plants	Reptiles
CHELSA Precip, 10000m 1 Day	7.05×10^5	4.35×10^8	1.10×10^7	1.18×10^7	3.94×10^6	9.95×10^6	7.86×10^7	9.37×10^5

Technical and Science Advancements

Count of annotated species occurrence records

Annotation Run	Amphibians	Birds	Fishes	Invertebrates	Mammals	Other	Plants	Reptiles
CHELSA Precip, 10000m 1 Day	7.05×10^5	4.35×10^8	1.10×10^7	1.18×10^7	3.94×10^6	9.95×10^6	7.86×10^7	9.37×10^5
CHELSA Precip, 10000m 15 Day	7.06×10^5	4.37×10^8	1.10×10^7	1.18×10^7	3.94×10^6	9.95×10^6	7.86×10^7	9.38×10^5
CHELSA Precip, 10000m 30 Day	7.07×10^5	4.39×10^8	1.10×10^7	1.18×10^7	3.94×10^6	9.95×10^6	7.86×10^7	9.40×10^5
CHELSA Precip, 1000m 1 Day	7.04×10^5	4.34×10^8	1.10×10^7	1.17×10^7	3.91×10^6	9.90×10^6	7.81×10^7	9.36×10^5
CHELSA Precip, 1000m 15 Day	7.05×10^5	4.35×10^8	1.10×10^7	1.17×10^7	3.92×10^6	9.90×10^6	7.81×10^7	9.38×10^5
CHELSA Precip, 1000m 30 Day	7.06×10^5	4.37×10^8	1.10×10^7	1.17×10^7	3.92×10^6	9.91×10^6	7.81×10^7	9.39×10^5
CHELSA Precip, 5000m 1 Day	7.05×10^5	4.35×10^8	1.10×10^7	1.18×10^7	3.94×10^6	9.95×10^6	7.86×10^7	9.37×10^5
CHELSA Precip, 5000m 15 Day	7.06×10^5	4.37×10^8	1.10×10^7	1.18×10^7	3.94×10^6	9.95×10^6	7.86×10^7	9.38×10^5
CHELSA Precip, 5000m 30 Day	7.07×10^5	4.39×10^8	1.10×10^7	1.18×10^7	3.94×10^6	9.95×10^6	7.86×10^7	9.40×10^5
Landsat8 EVI, 100m 16 Day	1.98×10^5	2.03×10^8	1.16×10^6	2.67×10^6	8.13×10^5	1.85×10^6	9.17×10^6	3.09×10^5
Landsat8 EVI, 100m 32 Day	2.61×10^5	2.63×10^8	1.45×10^6	3.41×10^6	1.06×10^6	2.35×10^6	1.17×10^7	3.86×10^5
Landsat8 EVI, 250m 16 Day	2.15×10^5	2.15×10^8	1.23×10^6	2.88×10^6	8.77×10^5	1.99×10^6	9.85×10^6	3.26×10^5
Landsat8 EVI, 250m 32 Day	2.74×10^5	2.72×10^8	1.48×10^6	3.55×10^6	1.10×10^6	2.43×10^6	1.22×10^7	3.98×10^5
Landsat8 EVI, 30m 16 Day	1.79×10^5	1.86×10^8	1.06×10^6	2.31×10^6	7.07×10^5	1.58×10^6	7.79×10^6	2.89×10^5
Landsat8 EVI, 30m 32 Day	2.42×10^5	2.48×10^8	1.41×10^6	3.05×10^6	9.47×10^5	2.06×10^6	1.03×10^7	3.69×10^5
Landsat8 EVI, 500m 16 Day	2.30×10^5	2.26×10^8	1.28×10^6	3.08×10^6	9.35×10^5	2.11×10^6	1.05×10^7	3.42×10^5
Landsat8 EVI, 500m 32 Day	2.85×10^5	2.79×10^8	1.50×10^6	3.67×10^6	1.14×10^6	2.51×10^6	1.26×10^7	4.08×10^5
MODIS EVI, 1000m 1 Day	1.95×10^5	1.91×10^8	1.83×10^6	3.62×10^6	1.27×10^6	2.17×10^6	1.52×10^7	2.80×10^5
MODIS EVI, 1000m 30 Day	5.66×10^5	4.33×10^8	7.35×10^6	7.46×10^6	2.90×10^6	5.27×10^6	4.07×10^7	7.33×10^5
MODIS EVI, 250m 1 Day	1.72×10^5	1.62×10^8	1.50×10^6	2.52×10^6	1.13×10^6	1.55×10^6	1.21×10^7	3.26×10^5
MODIS EVI, 250m 30 Day	4.11×10^5	3.53×10^8	7.08×10^6	5.81×10^6	2.17×10^6	4.29×10^6	3.66×10^7	6.30×10^5
MODIS EVI, 250m 90 Day	4.02×10^5	3.49×10^8	3.46×10^4	6.25×10^5	2.25×10^6	7.20×10^5	2.25×10^6	4.79×10^5
MODIS LST Day, 1000m 1 Day	1.06×10^5	1.38×10^8	2.67×10^5	1.95×10^6	5.93×10^5	9.18×10^5	6.23×10^6	1.84×10^5
MODIS LST Day, 1000m 30 Day	3.19×10^5	3.28×10^8	6.19×10^5	3.85×10^6	1.51×10^6	2.08×10^6	1.48×10^7	4.67×10^5
MODIS LST Day, 1000m 90 Day	1.16×10^5	1.43×10^8	1.16×10^4	2.03×10^5	5.40×10^5	1.94×10^5	5.56×10^5	2.06×10^5
MODIS LST Night, 1000m 1 Day	1.17×10^5	1.44×10^8	2.71×10^5	2.11×10^6	6.54×10^5	1.02×10^6	6.88×10^6	1.87×10^5
MODIS LST Night, 1000m 30 Day	3.13×10^5	3.27×10^8	6.12×10^5	3.84×10^6	1.50×10^6	2.06×10^6	1.48×10^7	4.81×10^5
MODIS LST Night, 1000m 90 Day	1.30×10^5	1.50×10^8	1.45×10^4	3.01×10^5	6.88×10^5	2.70×10^5	7.60×10^5	2.09×10^5
MODIS NDVI, 1000m 1 Day	1.95×10^5	1.91×10^8	1.83×10^6	3.62×10^6	1.27×10^6	2.17×10^6	1.52×10^7	2.80×10^5
MODIS NDVI, 1000m 30 Day	5.66×10^5	4.33×10^8	7.35×10^6	7.46×10^6	2.90×10^6	5.27×10^6	4.07×10^7	7.33×10^5
MODIS NDVI, 250m 1 Day	1.72×10^5	1.62×10^8	1.50×10^6	2.52×10^6	1.13×10^6	1.55×10^6	1.21×10^7	3.26×10^5
MODIS NDVI, 250m 30 Day	4.11×10^5	3.53×10^8	7.08×10^6	5.81×10^6	2.17×10^6	4.29×10^6	3.66×10^7	6.30×10^5
MODIS NDVI, 250m 90 Day	4.02×10^5	3.49×10^8	3.46×10^4	6.25×10^5	2.25×10^6	7.20×10^5	2.25×10^6	4.79×10^5

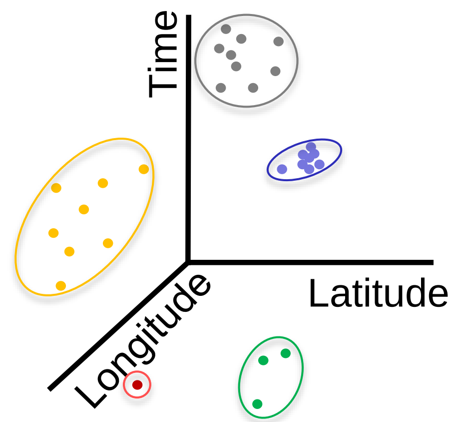
ESTO
Earth Science Technology Office

Technical and Science Advancements

Benchmarking results for key products + buffers: 10,000 records

Source	Product	Dataset	Spatial Buffer	Temporal Buffer	Mean Elapsed
Landsat 8	EVI	Global	30m	16d	0:44:00

Spatiotemporal clustering to increase computational efficiency



- Temporal buffers drive total compute time
- Comparable performance across products, faster for static layers

Technical and Science Advancements

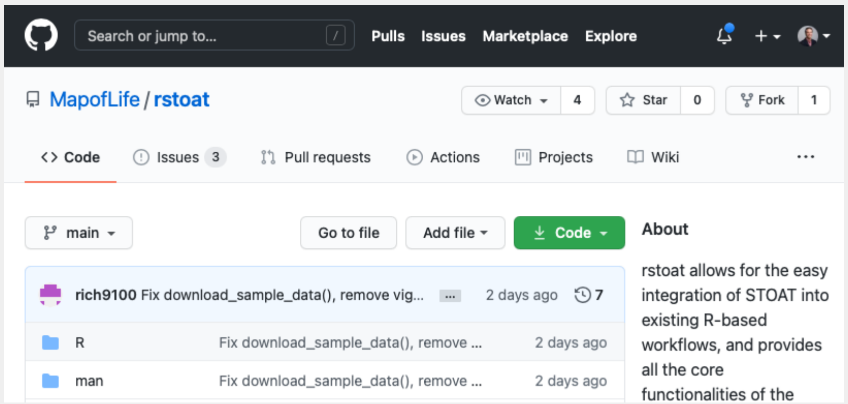
Benchmarking results for key products + buffers: 10,000 records

Source	Product	Dataset	Spatial Buffer	Temporal Buffer	Mean Elapsed
Landsat 8	EVI	Global	30m	16d	0:44:00
Landsat 8	EVI	Global	250m	16d	0:40:30
Landsat 8	EVI	Continental	30m	16d	0:36:00
Landsat 8	EVI	Local	30m	16d	0:22:00
MODIS	LST (Day)	Global	1000m	1d	0:22:00
MODIS	LST (Day)	Global	1000m	30d	0:59:00
MODIS	LST (Day)	Continental	1000m	1d	0:20:00
MODIS	LST (Day)	Local	1000m	1d	0:20:00
ESA CCI	Land Cover	Global	300m	365d	0:15:30
ESA CCI	Land Cover	Global	5000m	365d	0:14:00
ESA CCI	Land Cover	Continental	300m	365d	0:13:00
ESA CCI	Land Cover	Local	300m	365d	0:13:30
TNC Global	Human Modification	Global	1000m	Static	0:14:30
TNC Global	Human Modification	Global	10000m	Static	0:12:00
TNC Global	Human Modification	Continental	1000m	Static	0:12:30
TNC Global	Human Modification	Local	1000m	Static	0:13:00

- Temporal buffers drive total compute time
- Comparable performance across products, faster for static layers

Technical and Science Advancements

R Library



Stoat R library (v0.2) Available now at github.com/MapofLife/rstoat with BSD3 license

open source initiative Approved License

Extended documentation and examples available in Vignette

ESTO Earth Science Technology Office

Technical and Science Advancements

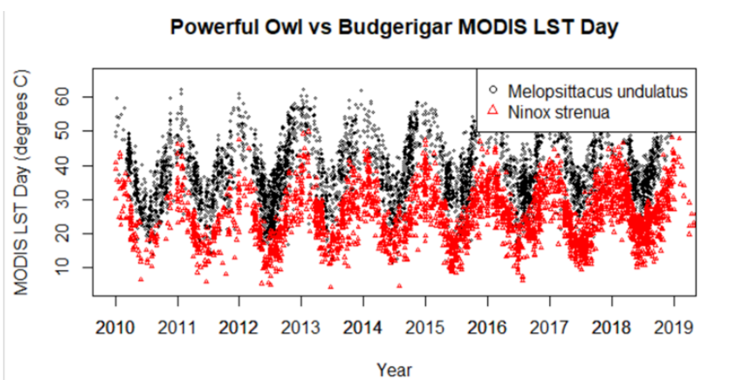
R package - Batch annotator

```
> #start_annotation_batch("dataset_id", "Annotation Title", "Layer Code(s)")
> start_annotation_batch(dataset_list$dataset_id[1], "powerful_owl_vignette",
c("modis-lst_day-1000-1", "modis-lst_day-1000-30"))
> start_annotation_batch(dataset_list$dataset_id[2], "budgerigar_vignette",
"modis-lst_day-1000-1")
```

#1
Specify Annotation occurrences for one or more species

```
> ninox_result_dir <- download_annotation(job_list$annotation_id[1])
Created directory: annotation_results
trying URL
'https://storage.googleapis.com/annotate/users/MOL_USER_56/9db8cc40-6502-4788-
894c-5afe89efa65b.zip?
Expires=1612319392&GoogleAccessId=206753940765%40project.gserviceaccount.com&S
ignature=...'
Content type 'application/octet-stream' length 2664965 bytes (2.5 MB)
downloaded 2.5 MB
```

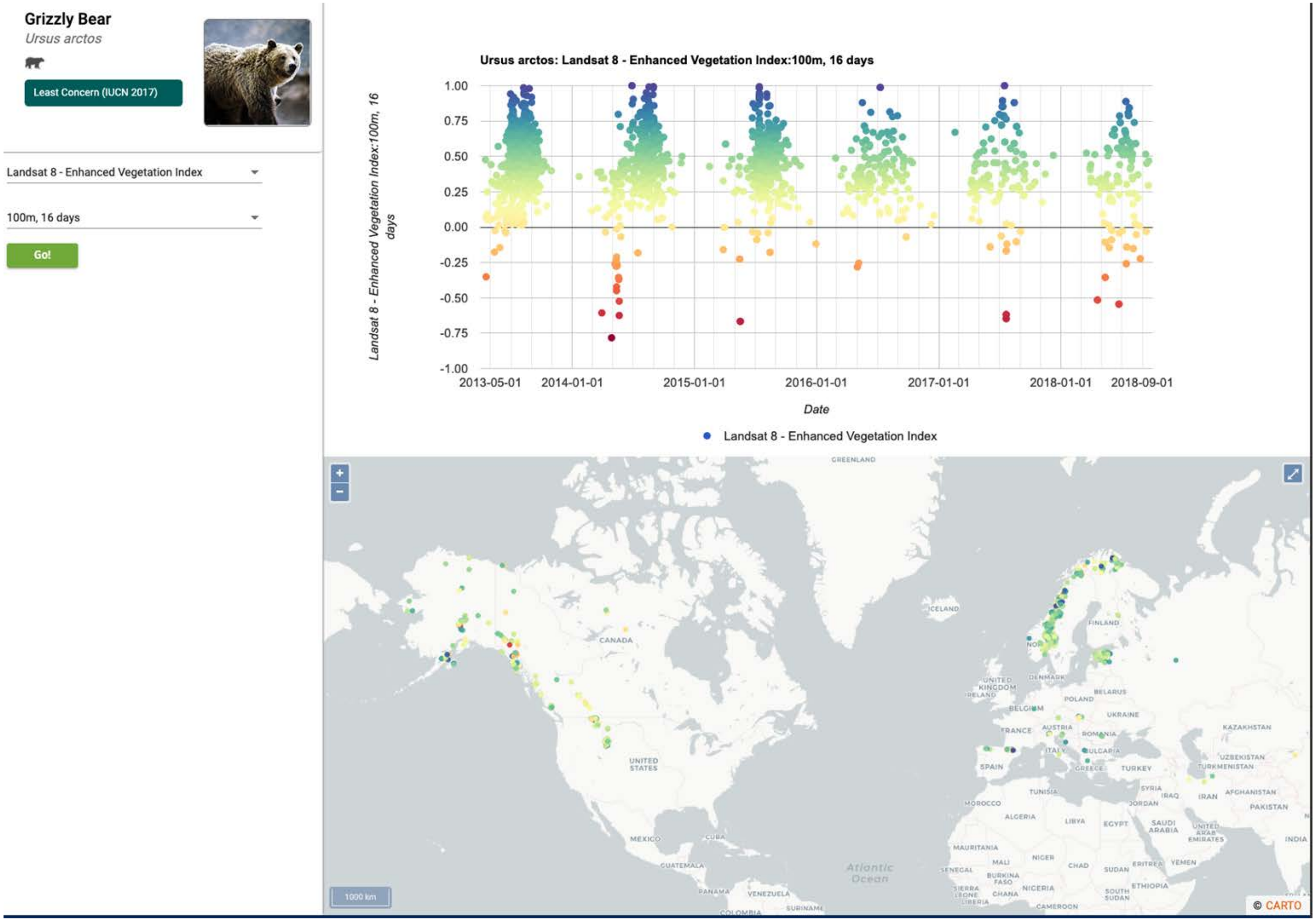
#2
Download annotated data as R data objects



#3
Use data in local analytical workflows in R

Technical and Science Advancements

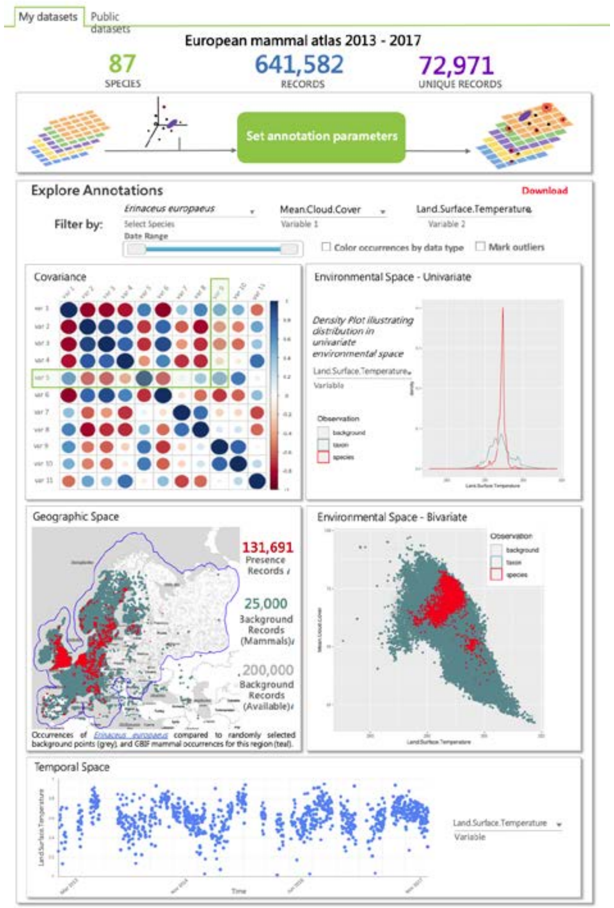
Driving an environmental niche tab in Map of Life



https://mol.org/species-dev/niche/Ursus_arctos

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Summary of Accomplishments and Future Plans



Future Plans

- System
  - Improve computational efficiency
  - Platform independence
- Web Interface
  - Improve user guidance for variable selection
- R package
  - Functions for effective variable selection

Summary of Accomplishments and Future Plans

- Soft-launch of Project at ESA 2020 virtual conference
  - Workshop: https://eco.confex.com/eco/2020/meetingapp.cgi/Session/17853
  - Poster: https://eco.confex.com/eco/2020/meetingapp.cgi/Paper/85131

- R-package pushed to CRAN
- STOAT Methods paper finished, ready for submission to Methods in Ecology and Evolution

Accomplished milestones:

- *Data connectors to the three platforms complete* *09/20*
- *Pre-annotations (for data up to that time) executed* *12/20*

Upcoming milestones:

- *Environmental data reports available on MOL* *03/21*
- *Final R library and documentation published on CRAN* *12/21*

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

Publications

A toolbox for environmental annotation of spatiotemporal biodiversity and ecological data. Richard Li, Ajay Ranipeta, John Wilshire, Jeremy Malczyk, Michelle Duong, Robert Guralnick, Adam Wilson, Walter Jetz

Journal : Methods in Ecology and Evolution

Submission: *submission Feb 2021*

A unifying framework for quantifying and comparing n-dimensional hypervolumes. Lu, M., K. Winner, and W. Jetz.

Journal : Methods in Ecology and Evolution

Submission: *in review*

Preprint: https://www.biorxiv.org/content/10.1101/2020.11.21.392696v2

Global cloud cover informed daily very high resolution downscaled land surface precipitation Dirk Nikolaus Karger, Adam Wilson, Colin Mahony, Niklaus E. Zimmermann, Walter Jetz

Target Journal: Nature Scientific Data

Submission: *submitted*

Preprint: *https://arxiv.org/abs/2012.10108*

List of Acronyms

Acronym	Definition
CMS	Content Management System.
CSV	Comma Separated Variables
DEM	Digital Elevation Model
DS	Decadal Survey
ESRI	Environmental Systems Research Institute
MOL	Map Of Life
GBIF	Global Biodiversity Information Facility
GCS	Google Cloud Storage
STOAT	Spatiotemporal Observation Annotation Tool
API	Application Programming Interface
QC	Quality Control

Thanks!

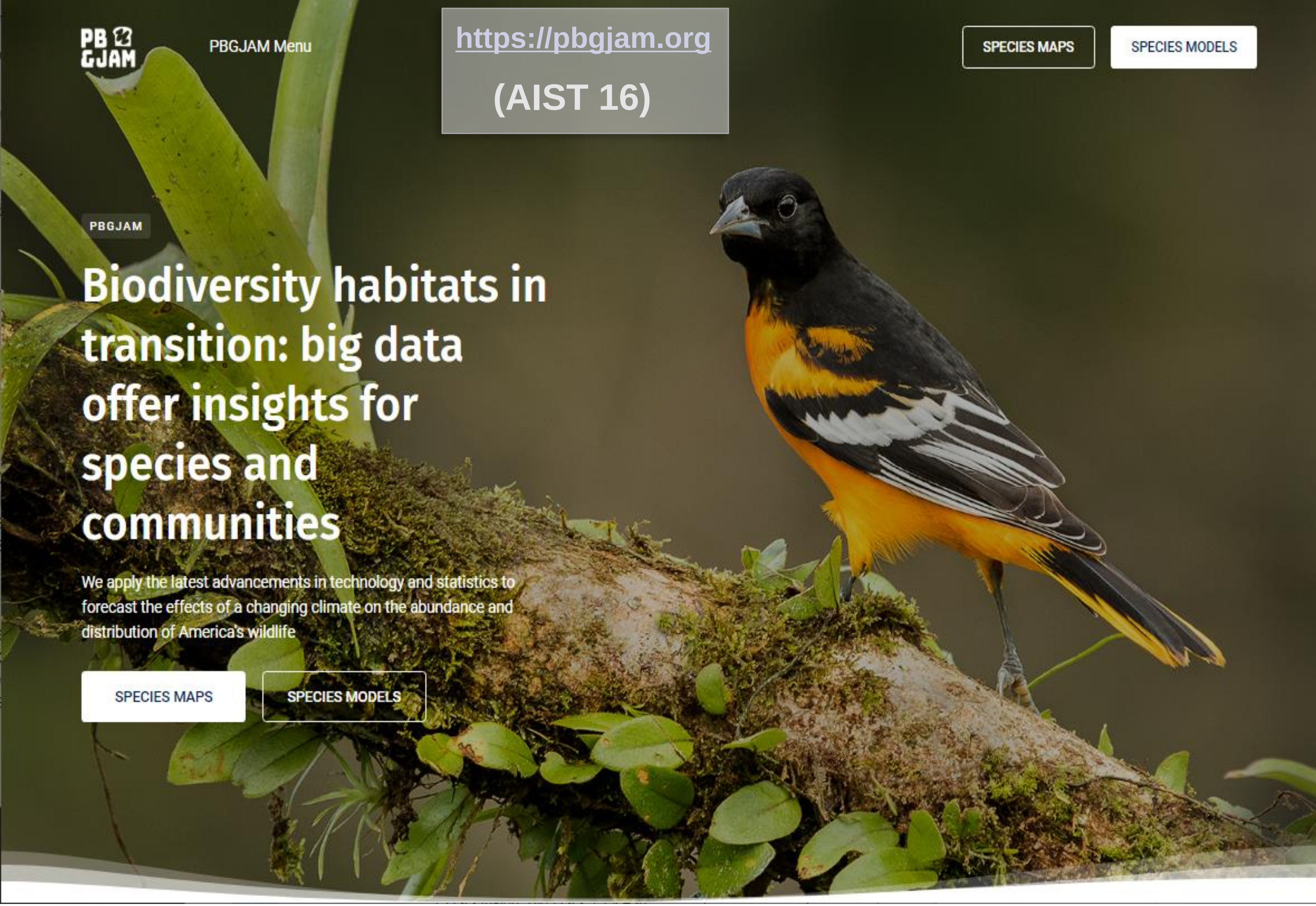
The bridge from canopy condition to continental scale biodiversity forecasts, including the rare species of greatest conservation concern

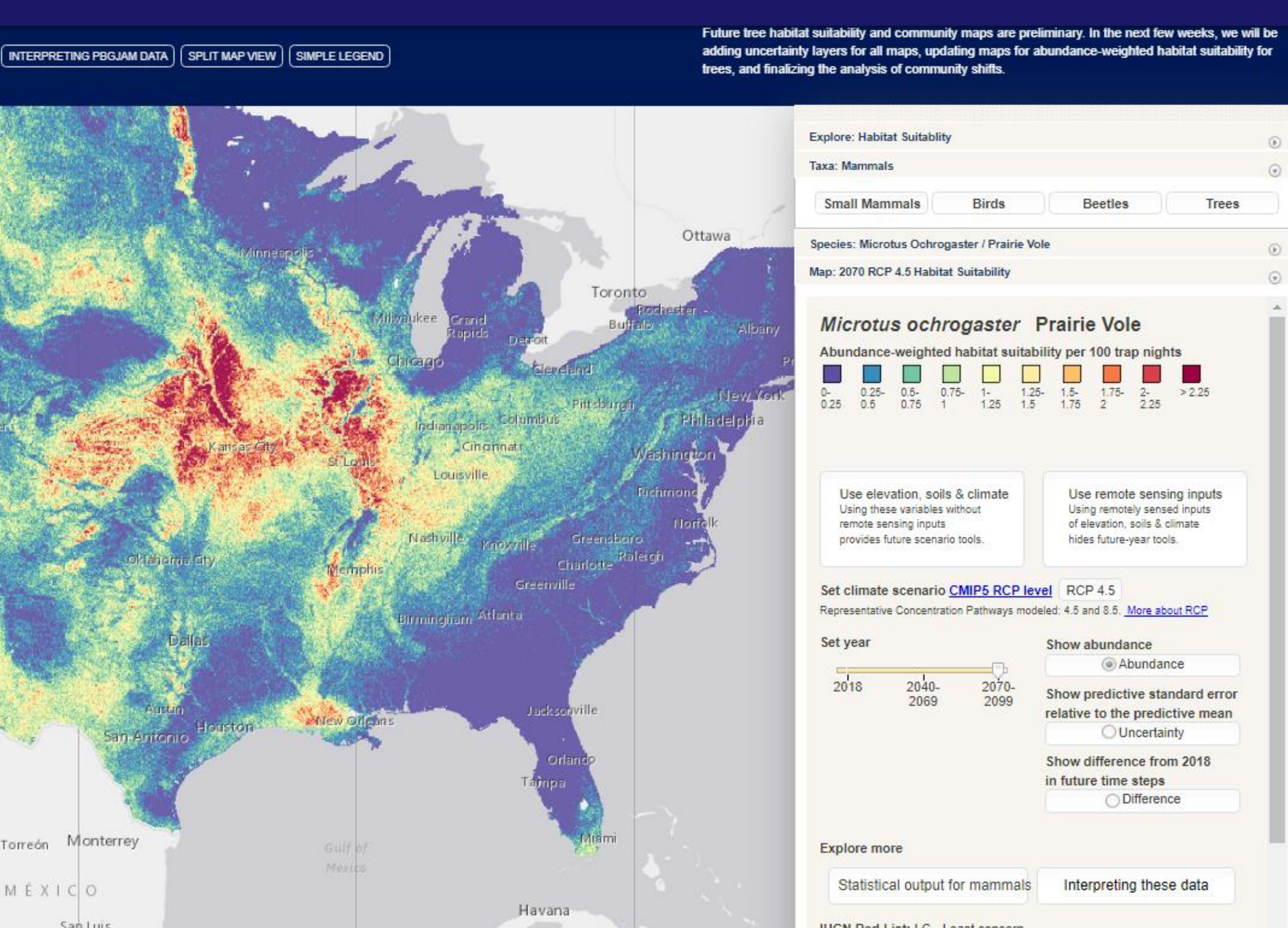
Jennifer Swenson, PI, Duke University
James Clark, Co-PI, Duke University

AIST-18-0063 Annual Group Technical Review
5 Feb 2021

https://pbgjam.org

(AIST 16)





Earth Science Technology Office

REGRESSION COEFFICIENTS | SENSITIVITY | COMMUNITIES | SENSITIVITY & REGRESSION COEF | ACCURACY

Input: Biodiversity data options

Small Mammals

Input: Environmental data options

climate + topography + soils

VIEW HABITAT SUITABILITY MAPS

Model outputs for Small Mammals

Mean posterior estimates: Fitted coefficients having 95% intervals that exclude zero. Use the dropdown menus below (1) to display regression coefficients for an explanatory variable in the model, (2) to color-code each species by a trait, or (3) to switch between common and scientific name. Environmental variables explain variation in species abundance across the map. Positive values ['Mean posterior estimate'] mean that higher values of a variable (e.g. high elevations) are associated with higher abundance, and vice versa.

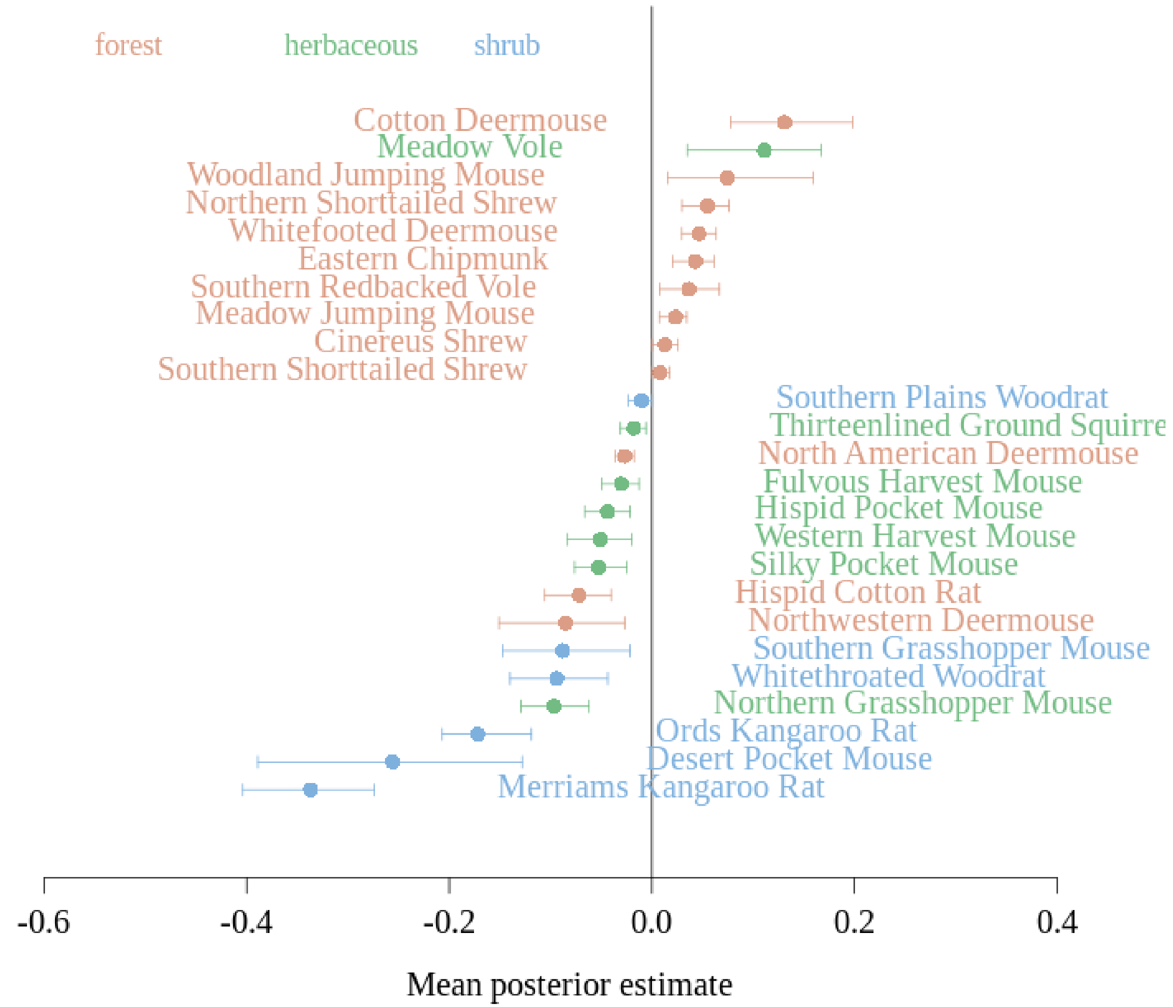
Environmental Variable

Apr-Sep precip-PET

Color-coded by trait

Habitat

To learn more about each trait click here



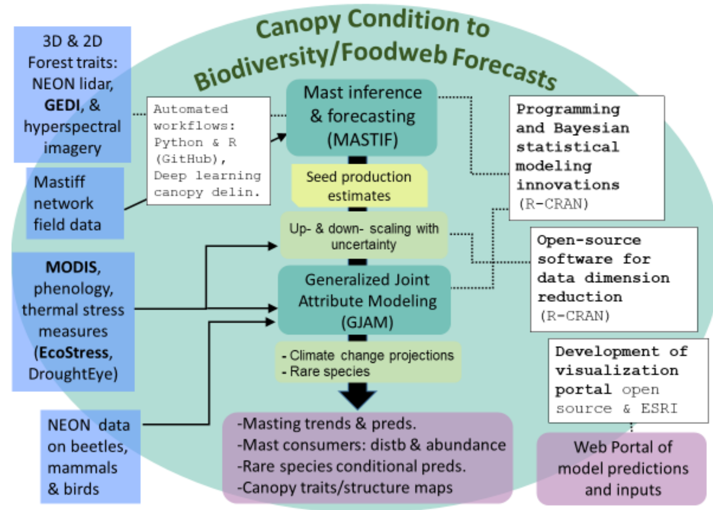
The Bridge from Canopy Condition to Continental Scale Biodiversity Forecasts, Including the Rare Species of Greatest Conservation Concern

PI: Jennifer Swenson, Duke University

Objective

Advance open-source modeling ensemble to solve scientific-technical challenges and develop an accessible user interface to access raw and processed data.

- Model the impacts of climate change on biodiversity with a focus on food energy fluctuations and thermal stress.
- Make predictions across the continental US and into the future under climate change for forest mast consumers: ground beetles, birds and small mammals.
- Target additional challenges of spatial scaling and conditional predictions for rare species.



Wire diagram: Canopy condition to continental scale biodiversity forecasts

Approach

Advancing open-source modeling ensemble to solve scientific-technical challenges and develop an accessible user interface to access raw and processed data:

- Update latest biodiversity surveys.
- Incorporate new statistical up- and down- scaling approach with non-smooth spatiotemporal dependence.
- Innovate conditional predictions for rare species via improvements to the generalized joint attribution model (GJAM) and mast inference and forecasting model (MASTIF).
- Develop biodiversity web portal to host field, predictor, and new modeled distributions that allow user interaction and multi-dimensional web visualizations.

Co-PI: J. Clark, Duke University

Key Milestones

• Data assimilation/preprocessing	11/20
• Canopy characterization linked to traits and reproductive effort in GJAM	05/21
• GJAM-MASTIF Model improvements	05/21
• Mast & consumer prediction local & national	11/21
• Full development of Biodiversity Portal	02/22

TRL$_{in}$ = 4 TRL$_{current}$ = 4

ESTO Earth Science Technology Office

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background: Ecological Forecasting

❖ **Technical framework**:

Automated workflows and probabilistic framework linking biodiversity and remote sensing for impacts of climate change on plants, ground beetles, birds and small mammals

❖ **Challenges**:

- Link remote sensing to high-dimensional canopy characteristics
- Predict species of conservation concerns, accommodating dependence structure

❖ **R&A and Application science goals**:

Biodiversity & Ecological Forecasting: *"factors that determine the distribution, abundance, movement, demographics, physical or genetic characteristics, behavior, and/or physiology of organisms on Earth …development of remote sensing tools, techniques, and associated models"*

Objectives:

Advance open-source modeling ensemble to solve scientific-technical challenges and develop an accessible user interface to access raw and processed data.

❖ Science objectives

- Evaluate climate change impacts on biodiversity, including species interactions and thermal stress.
- Predict efforts on mast consumers: ground beetles, birds and small mammals.

❖ Technical objectives

- Advance and integrate open source Bayesian analysis of environment-species interactions
- Develop data cube processing workflows for NEON observations and remotely sensed imagery
- Coherent, probabilistic inference and prediction of community change, including conditional prediction for rare species.
- Expand pbgjam.org site to a Biodiversity Portal with data, statistical, and cartographic visualizations.

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Technical Advancements Summary

- **Remote Sensing workflow:** 1-m 3-D canopy structure and characteristics (lidar + hyperspectral) for all NEON sites linked to land surface products from MODIS/VIIRS and soils;
- **gjamTime**: dynamic statistical workflow to quantify environment-species interactions
- **MASTIF:** dynamic inference food supply
- **Dimension reduction:** exploit spatio-temporal dependence in large arrays of remotely sensed canopy and climate
- **Conditional prediction:** rare species of conservation concern.

Science Advancements Summary

- **Continental-scale tree mast**
  - biogeographic fecundity and recruitment hotspots
  - fecundity change driven by indirect climate effects on growth
- **Tree Migration:** Tree migration is happening now, slowed in the East by fecundity, tracking climate in the West and North
- **Foliar traits and mast:** crown-level correlations between fecundity and foliar traits, collaboration with Phil Townsend.
- **Communities driven by environment-species interactions:** dynamic community responses in progress for birds and ground beetles to climate change

Technical Advancement:
Overview

Remote Sensing Input

NEON AOP

Hyperspectral foliar traits (with Townsend Lab)

Lidar 3D canopy structure and terrain attribtue

Other land surface products

Soil and Climate conditions

ECOSTRESS drought index

MODIS/VIIRS Phenology and Productivity

MASTIF

Pulsed seed and fruit prediction for North America

To date, **first** continental scale prediction

Biodiversity survey

Diverse Community (only NEON can provide such opporunity)

Vascular Plants

Ground Beetles

Birds

Small Mammals

GJAM/ GjamTime

Modeling framework

Probabilistic framework with **full uncertainty** for data, parameter, ,process, and prediction

Non-Gaussian for both space and time dimension

Direct interpretation on parameters and predictions

Multiple Data types and deal with zero inflation data

Conditional prediction for different communities (beetles and plants)

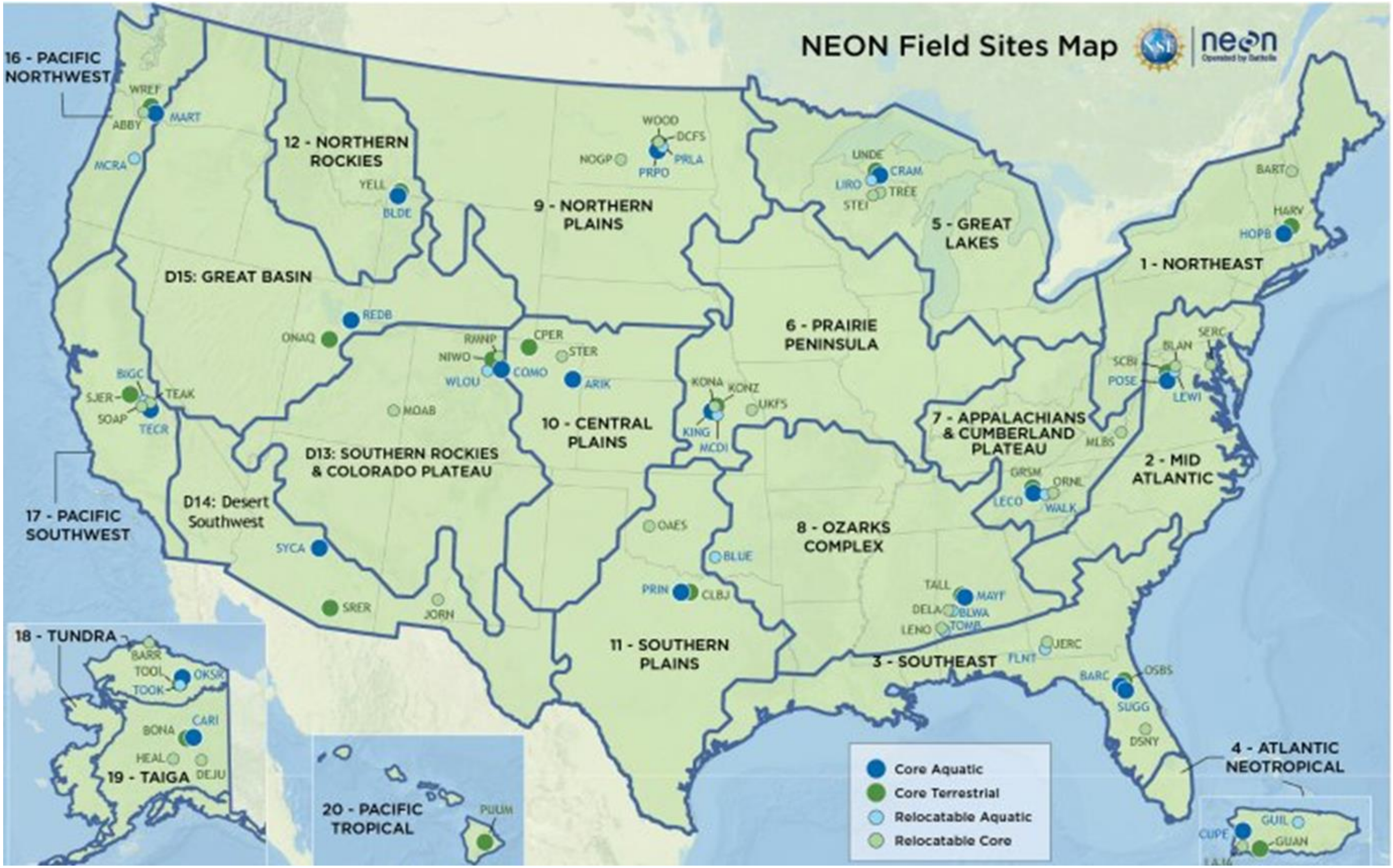
Multiple covariates at different scales

Dynamic model for species interaction inference and prediction

Open source

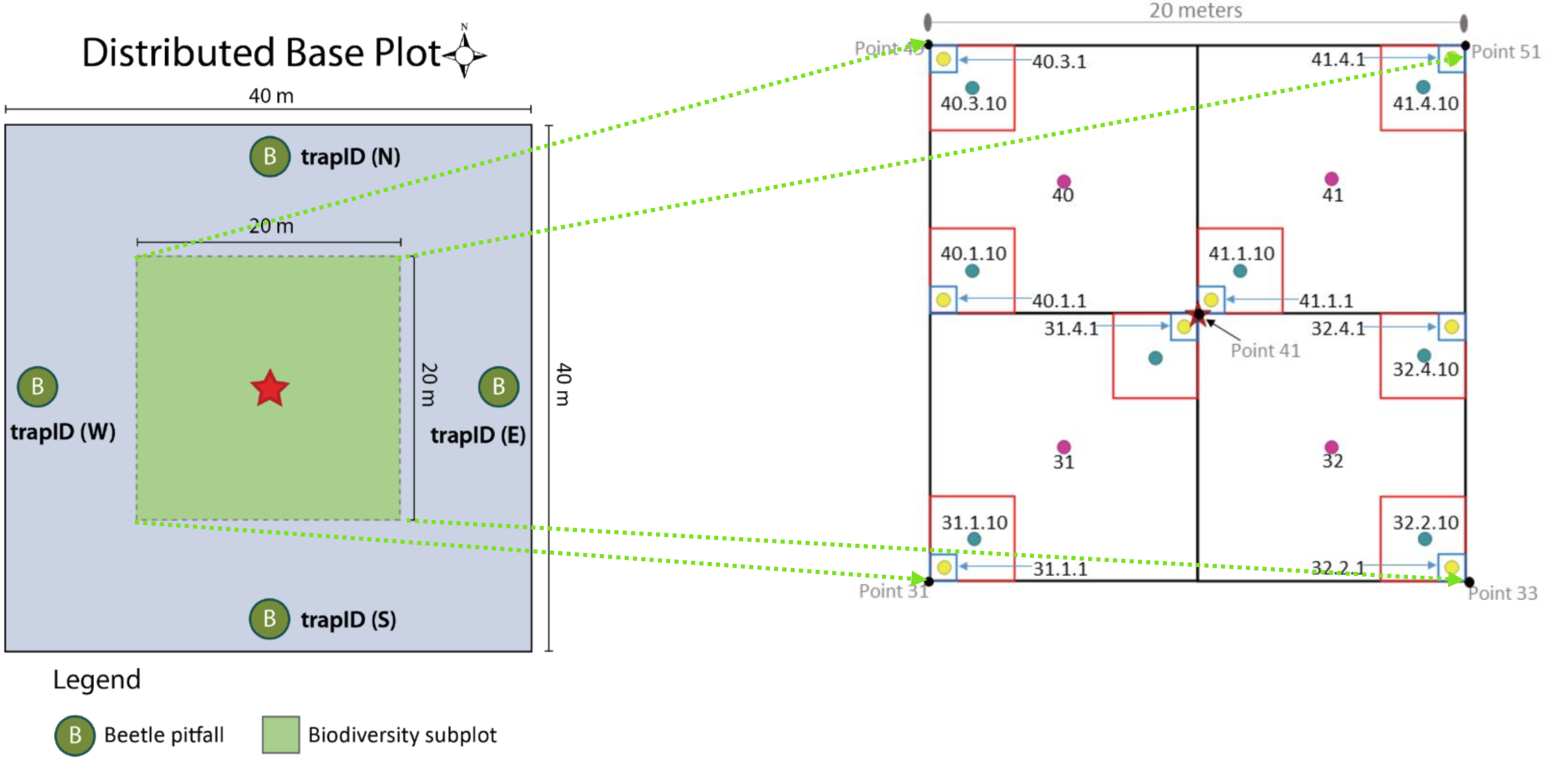
Technical Advancement:

NEON Comprehensive Biodiversity Data Collection



Technical Advancement:

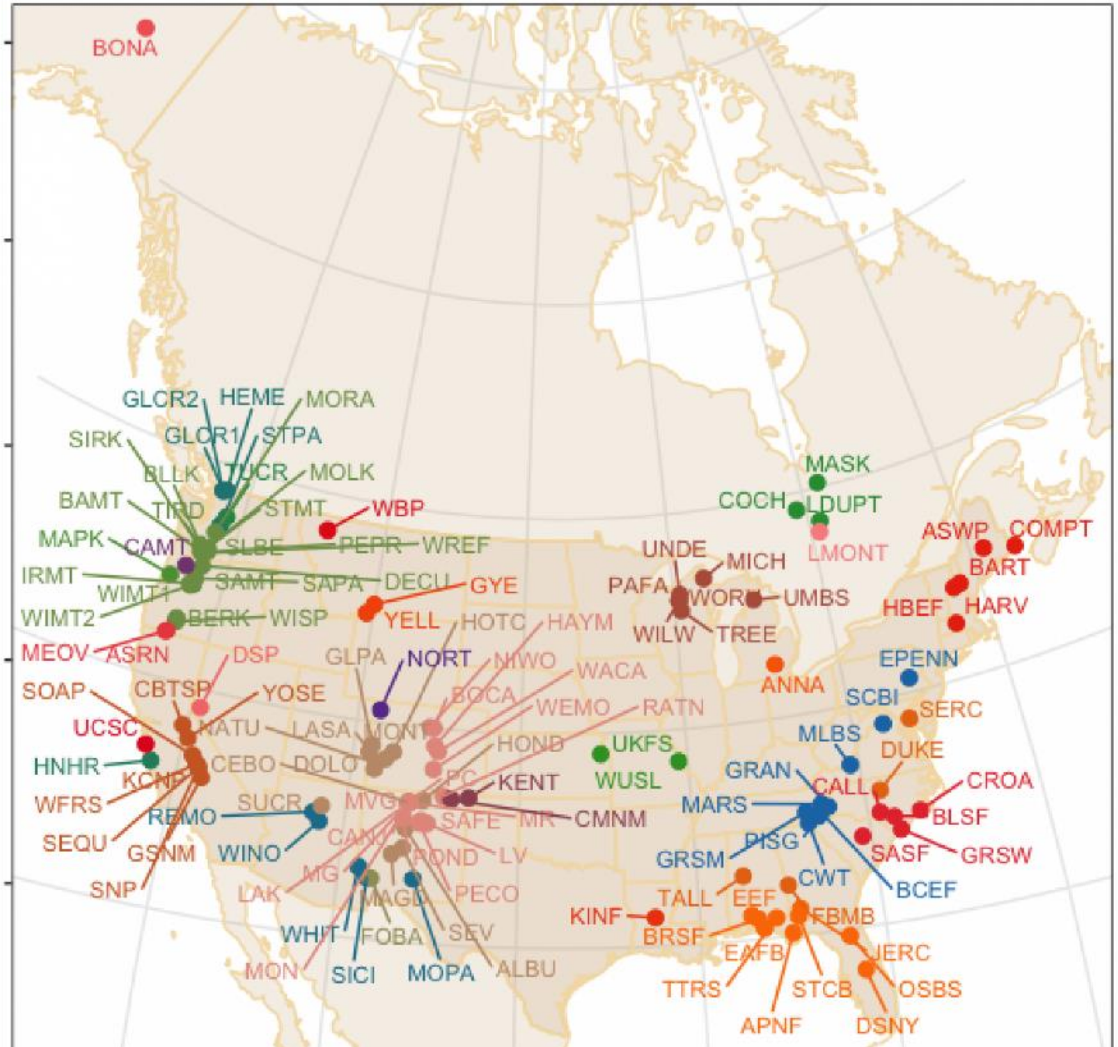
NEON Comprehensive Biodiversity Data Collection



NEON biotic data

- 14,000 observations @ Core Terrestrial Sites
- Consistent design
- Multiple species groups: plants, trees, ground beetles, birds, small mammals.

Technical Advancement:
Mast Inference and Forecast (MSTIF)



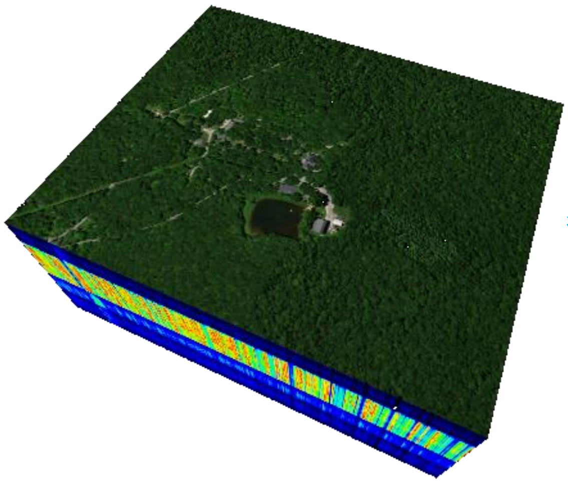
Long-term monitoring network plots and crop-count locations.

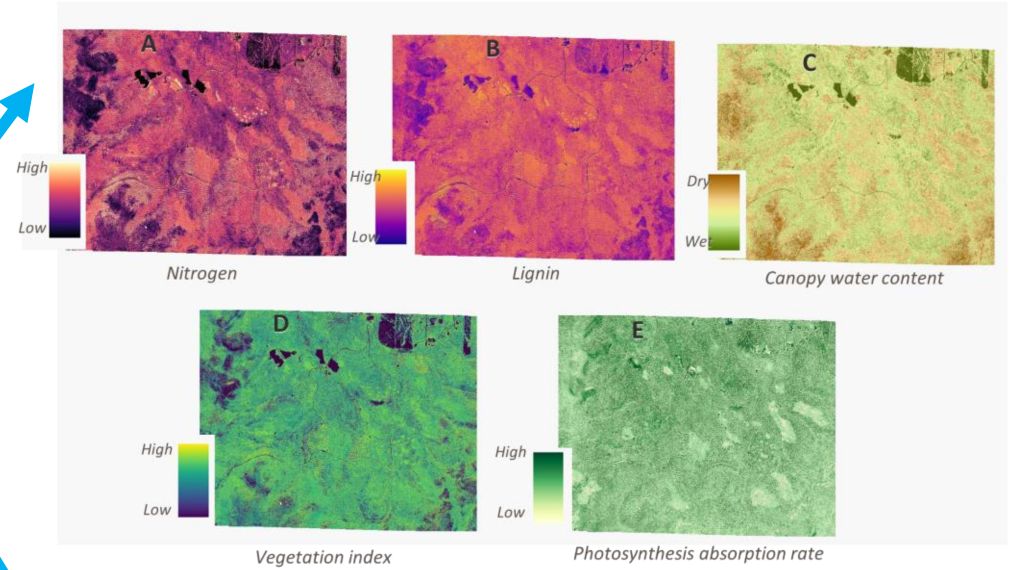
800 plots, 7M trees

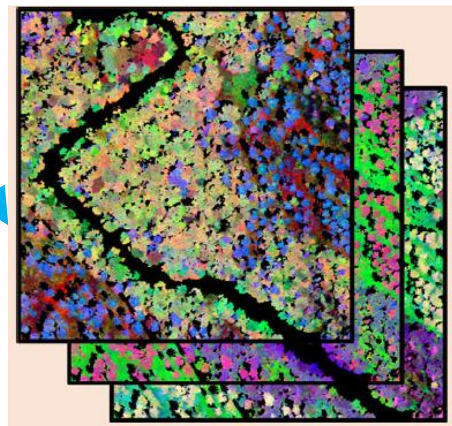
Base of the forest food web combines monitoring and opportunistic sampling

Technical Advancement:

Data integration & remote sensing workflow







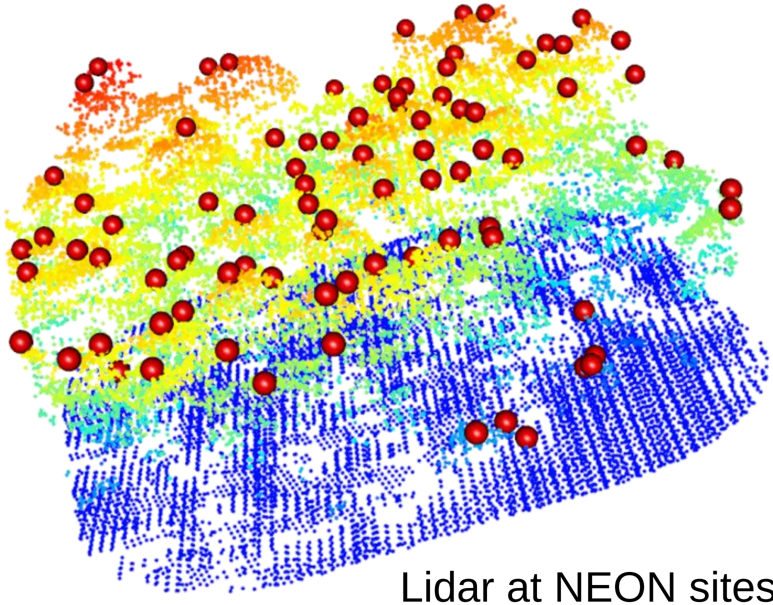
- LMA
- EWT
- SLA
- Carbon
- Nitrogen
- Cellulose
- Lignin
- Water
- NSC
- Chlorophylls
- Carotenoids
- Sugars
- Starches
- Phosphorus
- Potassium
- Calcium
- Sulfur
- …

Foliar traits for tree mast estimation

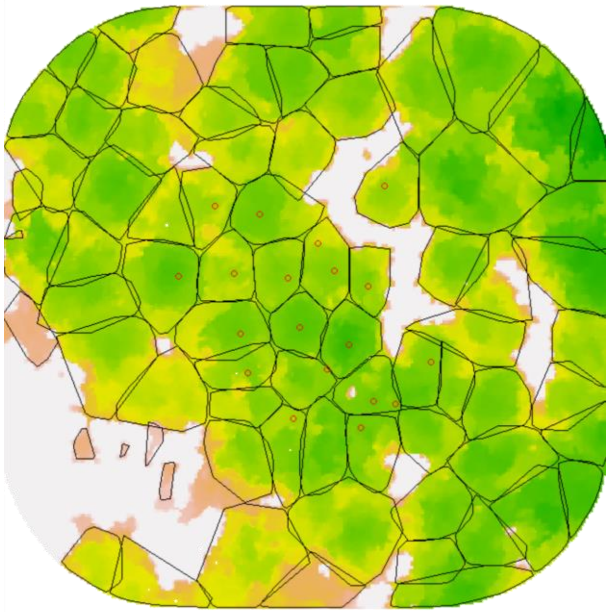
Collaboration with the Townsend lab

Spectral indices as canopy characteristics for biodiversity modeling

Technical Advancement:
Data integration & remote sensing workflow



Lidar at NEON sites



Forest canopy structure

Current: Beetles, plant, tree field plots
Ongoing: wall-to-wall mapping across entire NEON site

- Understory habitat structure
- Crown delineation for individual tree attributes

gjamTime

quantify and predict environment-species interactions

- *The emergent interactions that govern biodiversity change*, PNAS (Clark, et al. 2020)
- Code on CRAN: gjam 2.3.0
- Tutorial here: https://rpubs.com/jimclark/631209
- Vignettes: https://cran.r-project.org/web/packages/gjam/vignettes/gjamVignette.html

Technical Advancement:
New Statistical Approach

Modeling spatial bias in citizen science data

Citizen science databases like ebird and iNaturalist need bias-correction

Point-process to account for location and effort bias

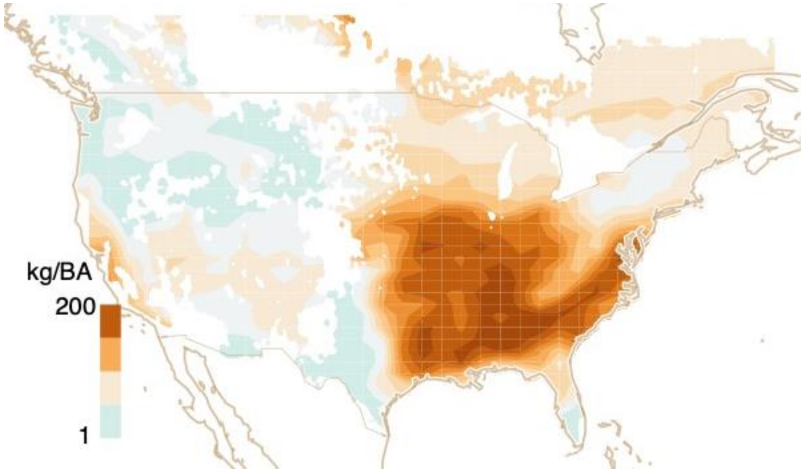
Novelty: will allow us to evaluate species distribution and abundance accounting for citizen science bias

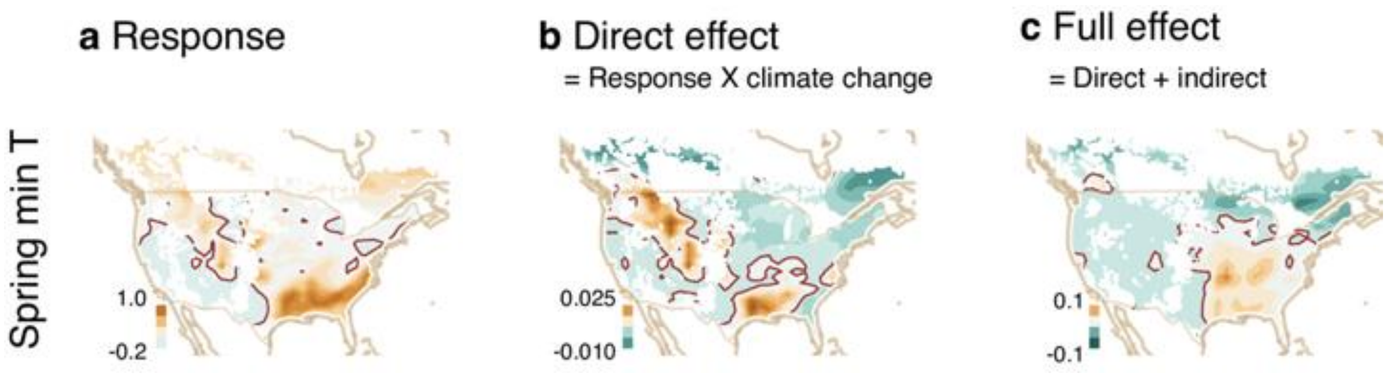
Tang et al., *in revision*, Modeling Spatially Biased Citizen Science Effort Through the eBird Database, EES, 2021

Science Advancements:
Continental-wide prediction of tree fecundity

- The biogeographic hotspot for fecundity
- Fecundity change driven by indirect climate effects: through tree growth and stand structure

Fecundity (seed mass/Basal Area)





Clark et al., in press ***Nature Communications***

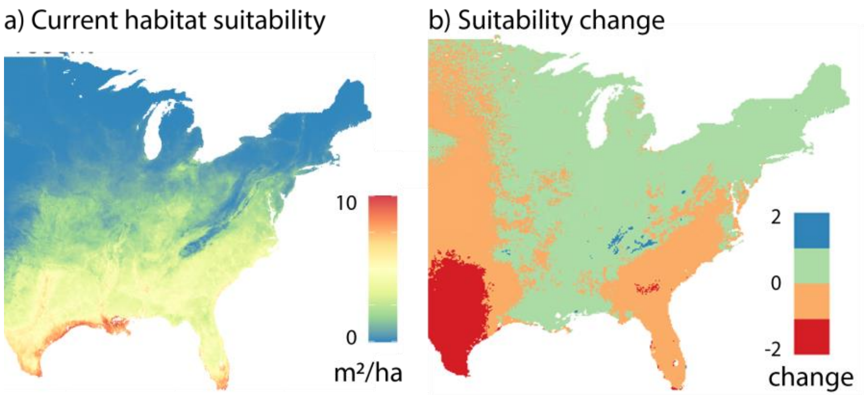
Science Advancements:
Migration study

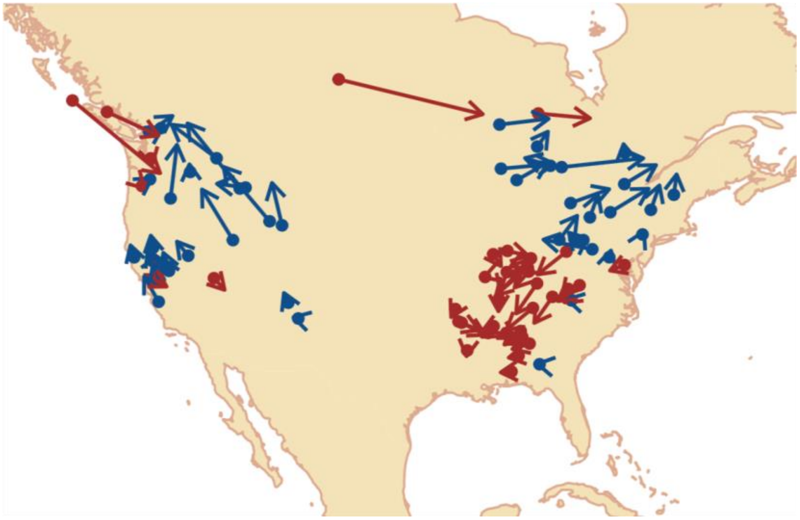
Can tree populations track climate change?

First continental scale synthesis of tree fecundity and recruitment rates

Species in the North and intermountain West are moving now.

Species in the Southeast are lagging



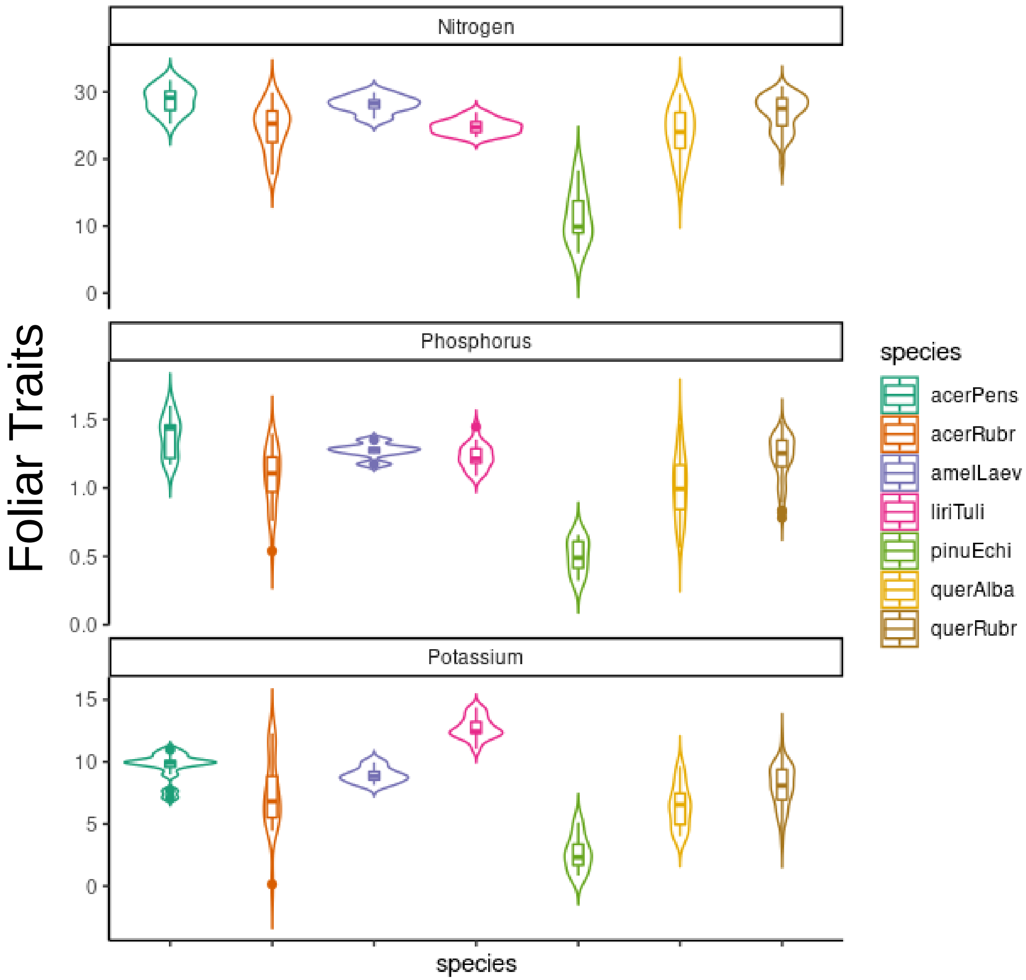


Sharma et al., in review

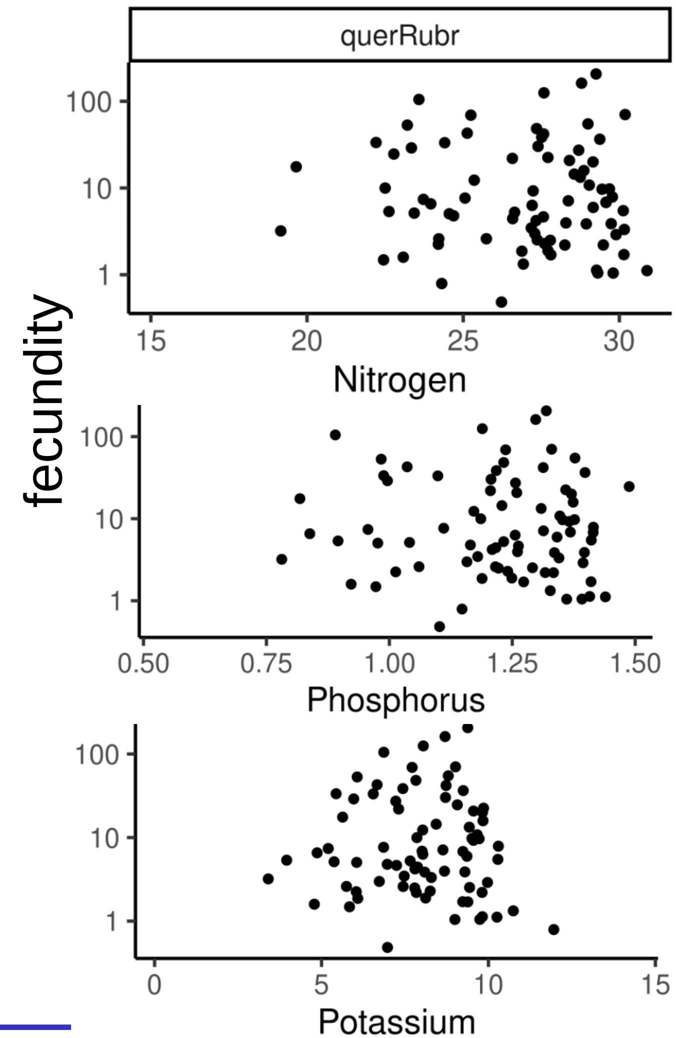
Science Advancement:

Tree mast and Foliar Traits at individual tree crown

Substantial differences in foliar traits across species; NEON Mountain Lake Biological Station



Weak correlation between foliar traits and fecundity (more sites to be added)

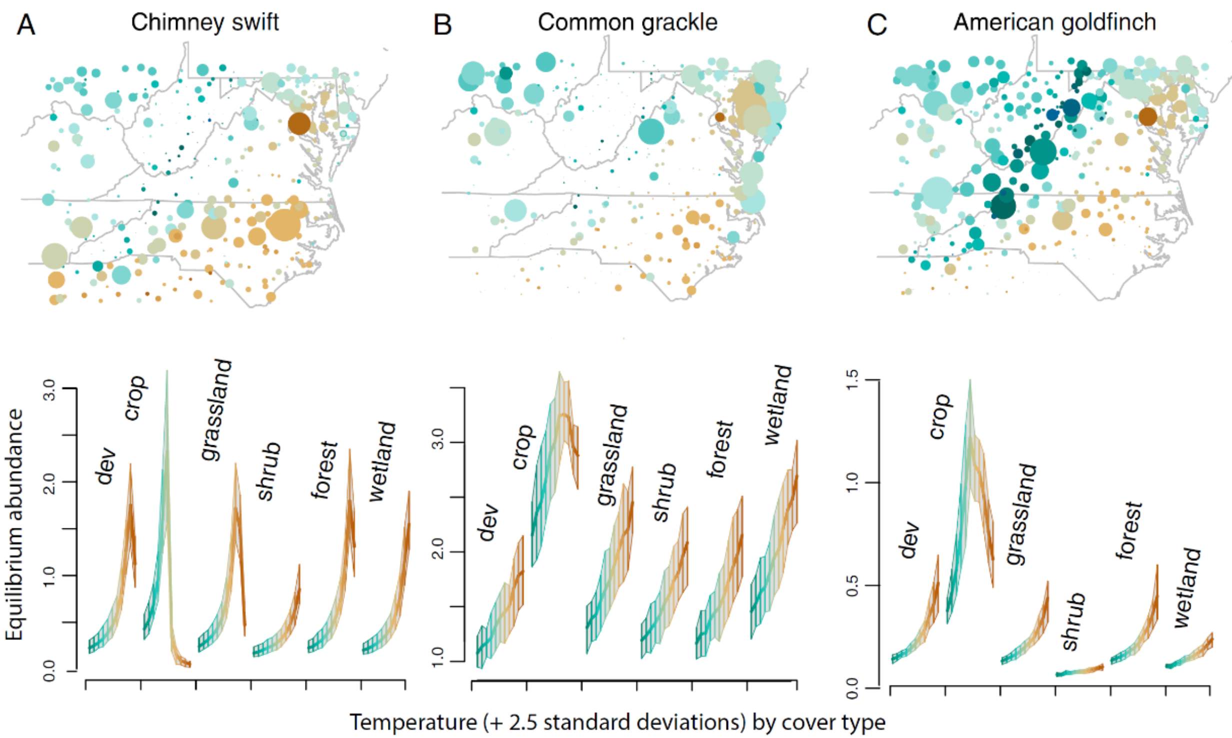


Science Advancement:

gjamTime captures environment species interaction

Environment-species interactions quantifiable with full uncertainty

Applications to determine impacts of climate change when species are responding to other species that are also responding to climate change

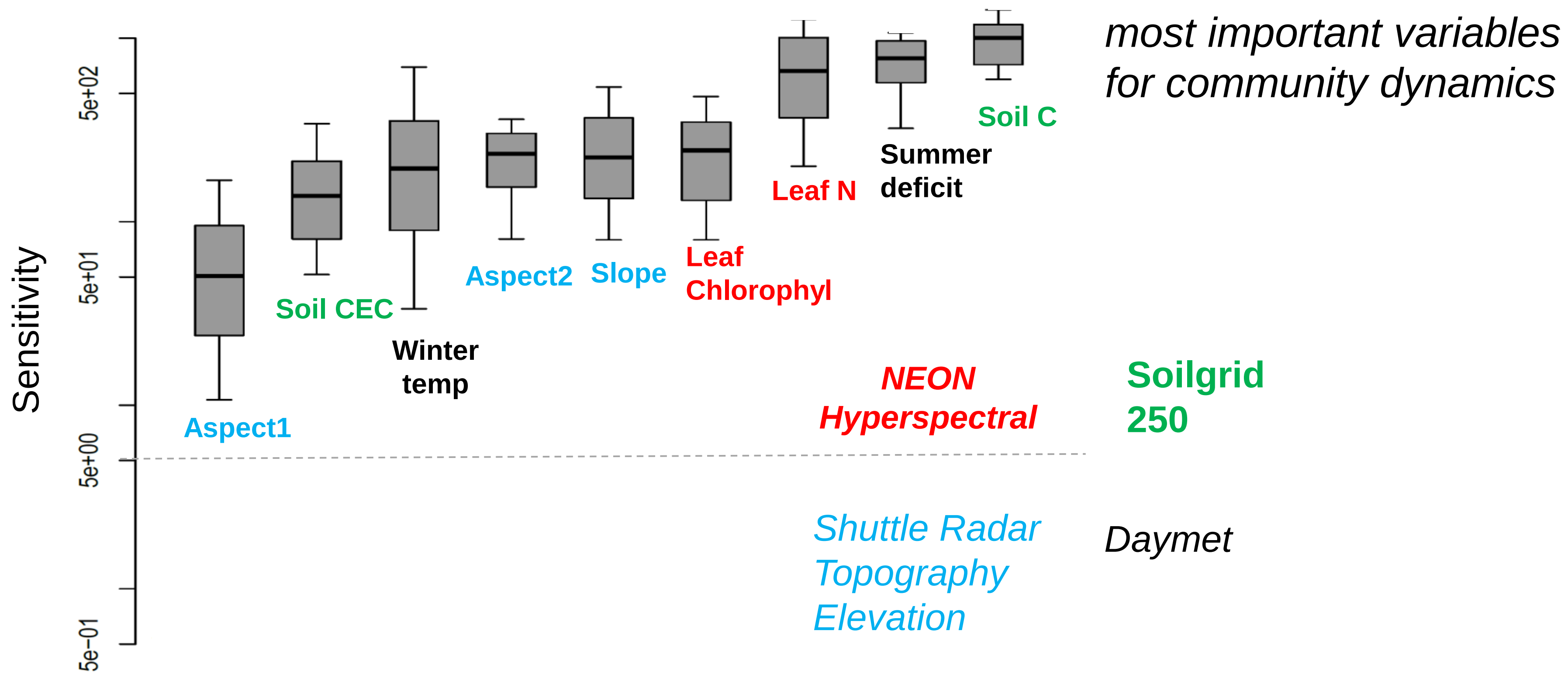


Response to temperature depends on land cover

Clark et al., PNAS 2020

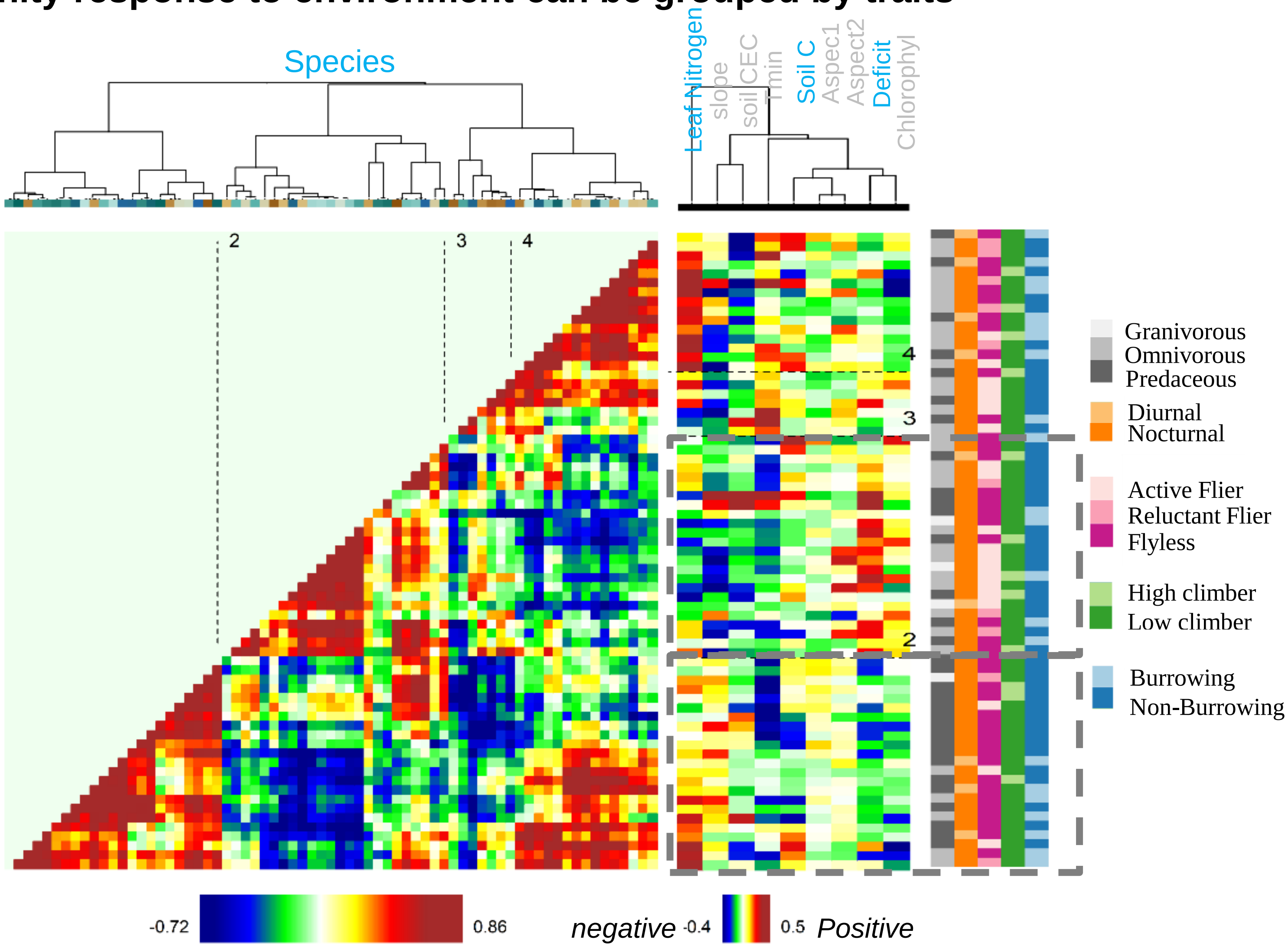
Science Advancement:
Technology application for biodiversity modeling

Beetles response to multi-scale environment variables



Science Advancement:
Technology application for biodiversity modeling

Community response to environment can be grouped by traits



ESTO Earth Science Technology Office

Science Advancement
Exploration of other NASA products

Continue to include/test NASA products in the mastif/gjam model:

→ MODIS and VIIRS Phenology to model tree mast

→ MODIS Land Surface Temperature to model ground beetle abundance

→ GEDI forest height model to estimate tree mast

→ ECOSTRESS for the drought stress western sites (e.g.SOAP)

Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Summary of Accomplishments

Technology development

Task 1: Environment-species interactions:

- ❖ Dynamic GJAM (gjamTime) applied continent-wide
- ❖ Spatial biases in citizen science data: ebird and BBS
- ❖ Conditional prediction: tested using ground beetles and vascular plants

Results: open-sources R packages

Task 2: Remote sensing workflow:

- v NEON AOP Lidar processing (Point cloud data to canopy characterization)
- v NEON AOP Hyperspectral data (1-m resolution)
  - v With Townsend lab re-preprocessing hyperspectral (BRDF, etc.)
- v Soil fertility from soilgrid250 (250-m spatial resolution)
- v Canopy segmentation with Lidar

Anticipated results: continental-wide workflow for multi-scale covariates

Technology application

Task 3: Biodiversity analysis

- ❖ Plants: crown delineation to link tree mast and foliar traits
- ❖ Beetles: dynamic response to climate change and habitat
- ❖ Birds: integrating BBS and ebird

Anticipated results: development of gjam/mastif food web

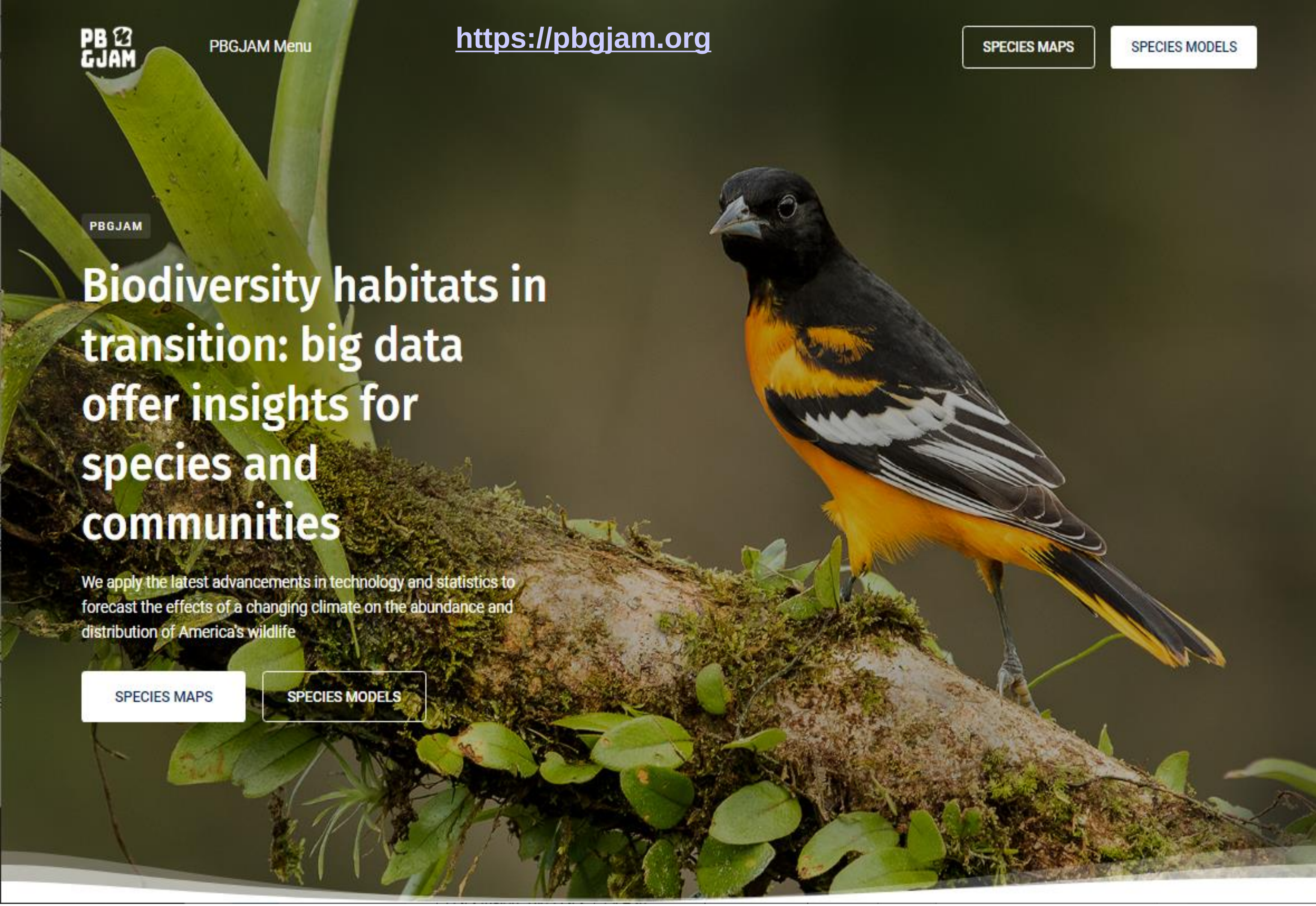
ESTO
Earth Science Technology Office

Future Plans

Next 6 months

- Remote sensing workflow:
  - Integration of Hyperspectral, lidar, MODIS/VIIRS land surface products, soil, terrain, etc. to modeling mast and biodiversity
  - ECOSTRESS testing for drought-stressed NEON sites
- Modeling: rare species, mast consumer prediction, up-/down-scaling with uncertainty
- Prediction of mast consumers: mammal, birds, rare species
- Website/portal: begin design & implementation

https://pbgjam.org



Presentation Contents

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

Publications

Publications

Clark, et al. In press. Continent-wide tree fecundity driven by indirect climate effects. ***Nature Communications***

Clark, J. S., C. L. Scher, and M. Swift. 2020. The emergent interactions that govern biodiversity change. ***Proceedings of the National Academy of Sciences***, 202003852, https://doi.org/10.1073/pnas.2003852117.

Qiu, T., C. Song, J. S. Clark, B. Seyednasrollah, and N. Rathnayaka. 2020. Understanding the continuous phenological development at daily time step with a Bayesian hierarchical space-time model: impacts of climate change and extreme weather events. ***Remote Sensing of Environment***, in press.

Conference Papers

IGARSS:

Community reorganization response to climate change: species interaction, state-space modeling, and food webs. Jennifer Swenson, Tong Qiu, Amanda Schwantes, Christopher Kilner, Chase Nunez, Lane Scher, Shubhi Sharma, James S. Clark

ESA:

Capturing emergent interactions that govern food web dynamics with climate change. Tong Qiu, C. Lane Scher, Margaret E. Swift, Jennifer J. Swenson and James S. Clark

Is citizen science a reliable source of big data? Identifying biases in eBird records. C. Lane Scher and James S. Clark

Understanding the diverse responses of South African savanna communities to climate change Margaret E. Swift, Steven I. Higgins and James S. Clark

North American tree migration paced by fecundity and recruitment through contrasting mechanisms east and west Shubhi Sharma et al.

Interactions that control the pace of forest change in North America. James S. Clark et al.

AGU:

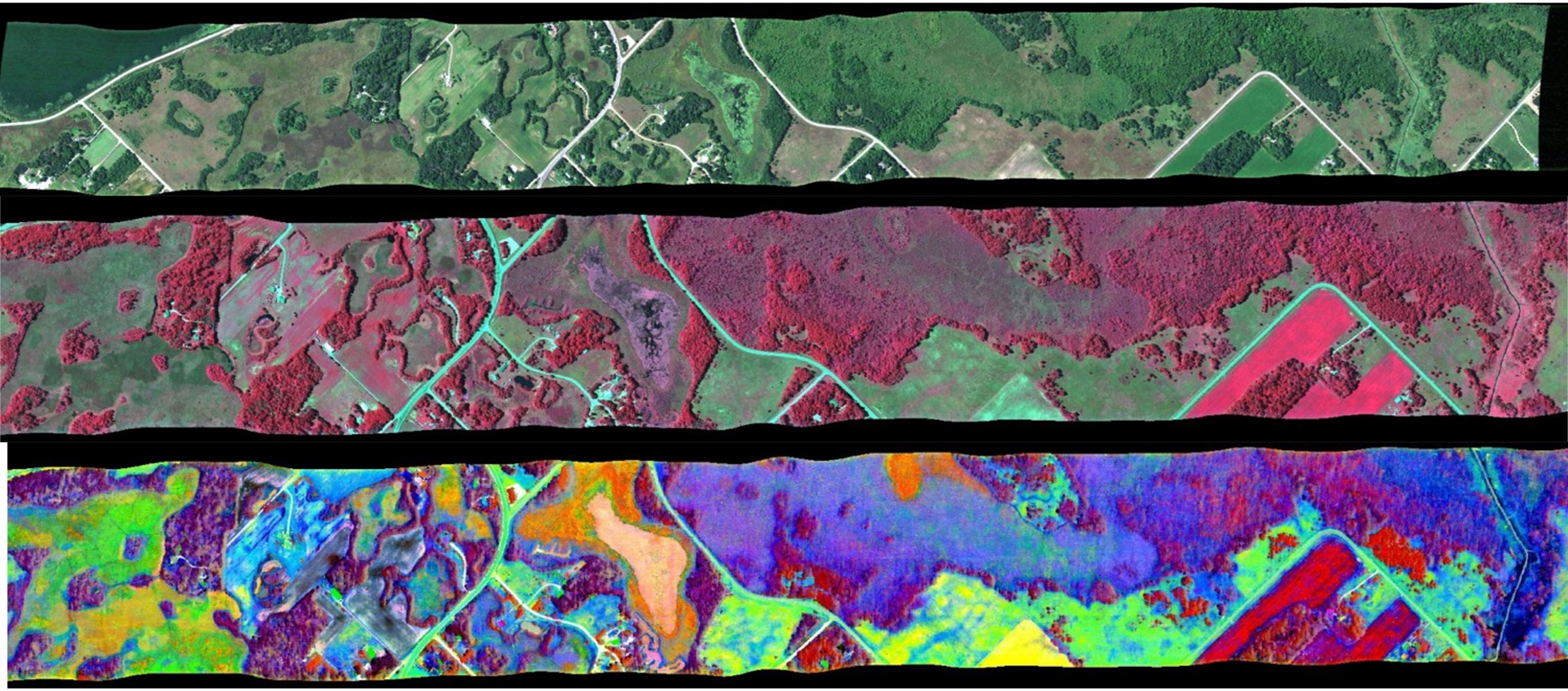
Dynamic responses of ground beetles to climate change and habitat characteristics Tong Qiu, Christopher Kilner, Jennifer Swenson, and James S. Clark

Acronyms

List of Acronyms

AOP Airborne observation platform
CRAN Comprehensive R Archive Network
GJAM Generalized Joint Attribute Model
MASTIF Masting inference network
NEON National Ecological Observatory Network

GeoSPEC (renamed ImgSPEC)



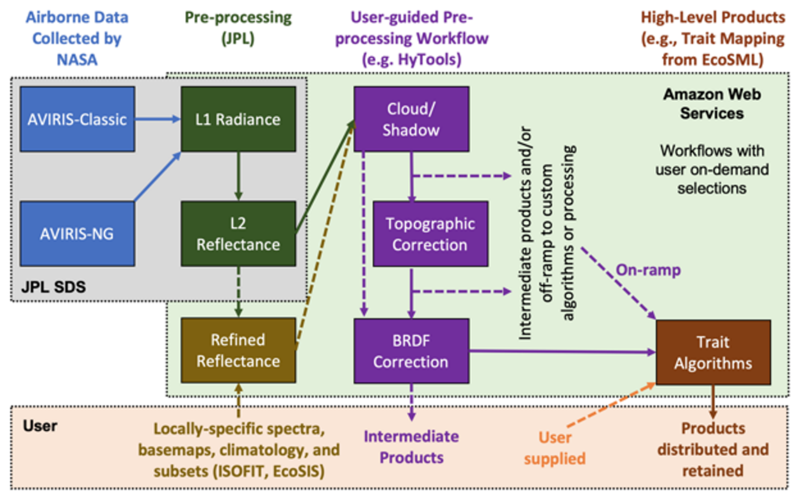
Philip Townsend (PI, U. Wisconsin, Madison)
Natasha Stavros (Co-I/JPL Lead exiting, JPL, Caltech)
Hook Hua (Co-I/JPL Lead entering, JPL, Caltech)
AIST-18-0043 Annual Review for February 5, 2020

On-Demand Geospatial Spectroscopy Processing Environment on the Cloud (GeoSPEC; renamed ImgSPEC)

PI: Philip Townsend, University of Wisconsin - Madison

Objectives

- Prototype a science data system to satisfy SBG's unprecedented data volume and user/algorithm diversity.
- **End-to-end on-demand cloud framework for imaging spectroscopy** Level 1 calibrated radiance through Level 3+ products including atmospheric and surface reflectance retrieval, BRDF correction, topographic correction, L3 retrieval algorithms, mosaicking, and analytics
- **Focus on SBG vegetation algorithms** in 2017 ESAS.
- Implement alternative open-source algorithms and user-defined parameterization
- Link to existing algorithm databases (EcoSML.org) and spectra (EcoSIS.org)
- Maintain provenance of on-demand user workflow with opportunity to modify as algorithms improve



Example workflow for vegetation traits. Solid lines: default workflow. Dashed lines: user-provided configuration options to improve corrections (e.g. L2 atmospheric/BRDF using spectra from EcoSIS in ISOFIT), implement alternative trait algorithms (e.g. from EcoSML), or eventually supply new algorithms.

Approach

- Leverage Earth/GIBS visualization and image mosaicking technology, Common Mapping Client (CMC) open-sourced software for predictive analytics
- Overcome download data volumes with Amazon Web Services platform accessing NASA AVIRIS-C and AVIRIS-NG spectra from JPL science data system
- Implement on-demand analysis with user on-ramps and off-ramps for alternative workflows with higher computational needs or customized parameterizations

Co-I's/Partners: Natasha Stavros, Hook Hua, Thomas Huang, Winston Olson-Duvall, David R. Thompson, George Chang, Sujen Shah, Justin Merz, William Phyo, Ryan Pavlick

Key Milestones

- ImgSPEC portal and connection to AVIRIS SDS 06/20
- Workflow testing on AWS with ISOFIT (atmospheric corr.) and HyTools (BRDF/topo corr. and veg. trait maps) 11/20
- Demonstrate full workflow including EcoSIS/EcoSML API 05/21
- Full demonstration of on-demand processing with on- and off-ramp options 10/21

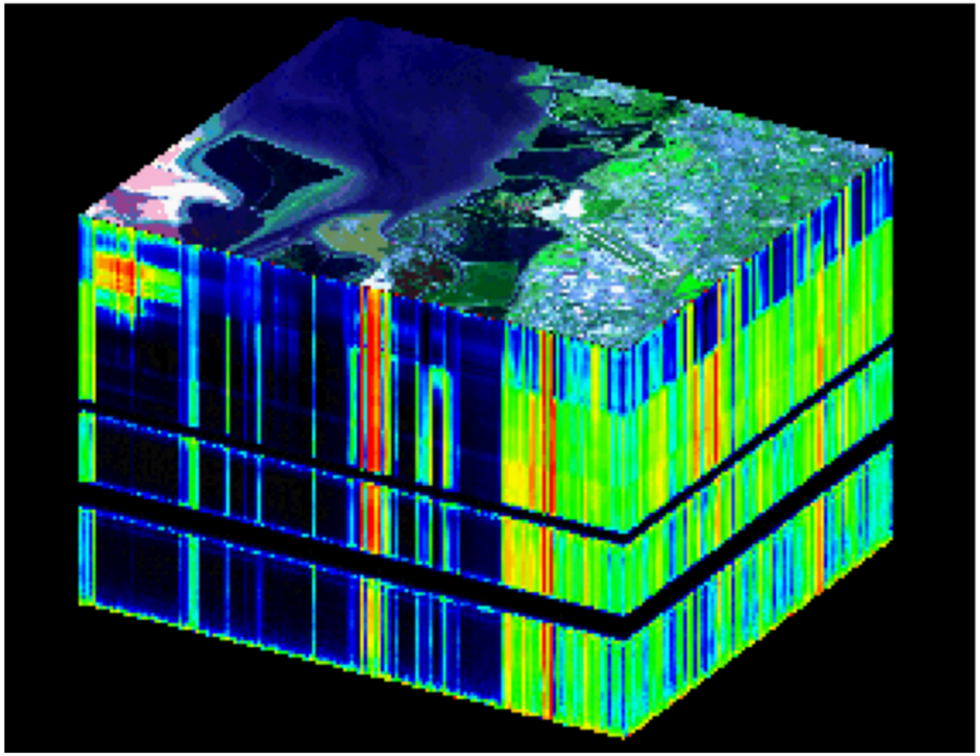
TRL_{in} = 3 **$TRL_{current}$= 4** **TRL_{out} = 5**

Presentation Contents

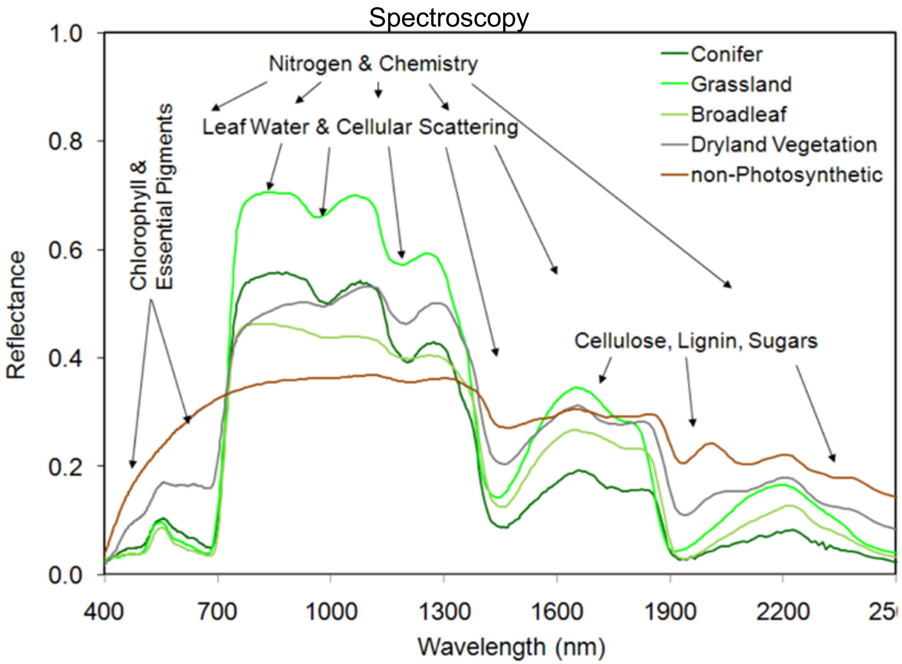
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background: SBG DO Study

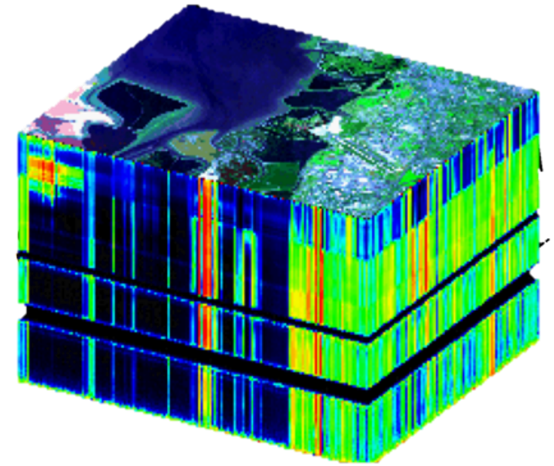
- The most recent Earth Science Decadal Survey was released by the National Academies in 2017
- 5 Designated Observables (DOs) – the most highly recommended Earth observables to change the current understanding of the Earth System
- Surface Biology and Geology (SBG) – open data access global imaging spectrometer at regular repeat observation

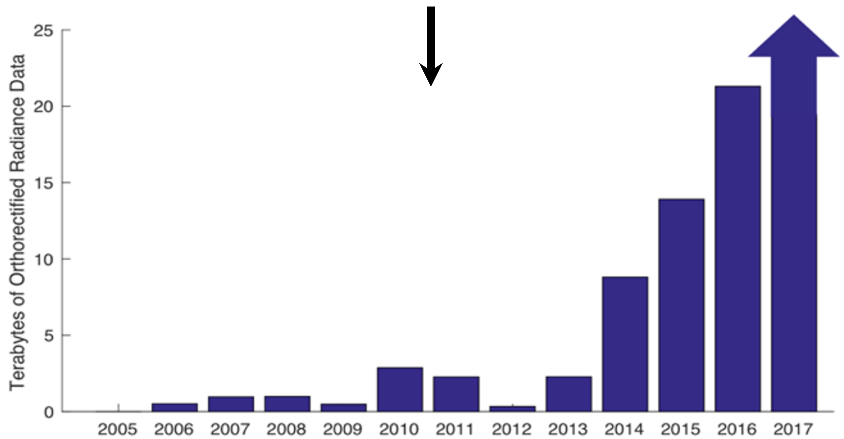


SBG

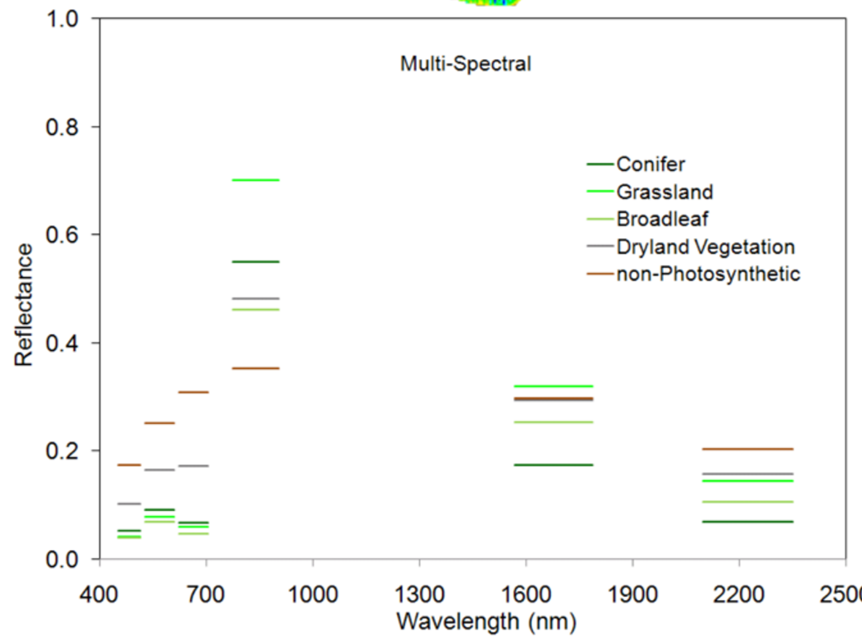


Spectroscopy
Nitrogen & Chemistry
Leaf Water & Cellular Scattering
Chlorophyll & Essential Pigments
Cellulose, Lignin, Sugars
Conifer
Grassland
Broadleaf
Dryland Vegetation
non-Photosynthetic
Reflectance
1.0
0.8
0.6
0.4
0.2
0.0
400
700
1000
1300
1600
1900
2200
Wavelength (nm)





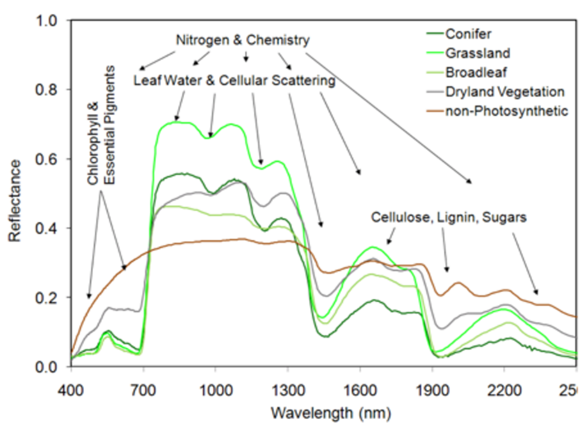
Terabytes of Orthorectified Radiance Data
25
20
15
10
5
0
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017



Multi-Spectral
Conifer
Grassland
Broadleaf
Dryland Vegetation
non-Photosynthetic
Reflectance
1.0
0.8
0.6
0.4
0.2
0.0
400
700
1000
1300
1600
1900
2200
2500
Wavelength (nm)

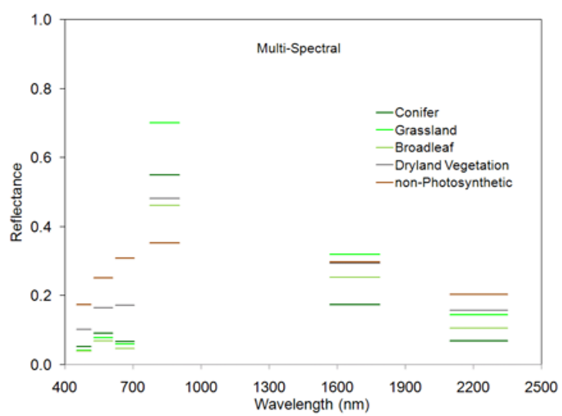
SBG

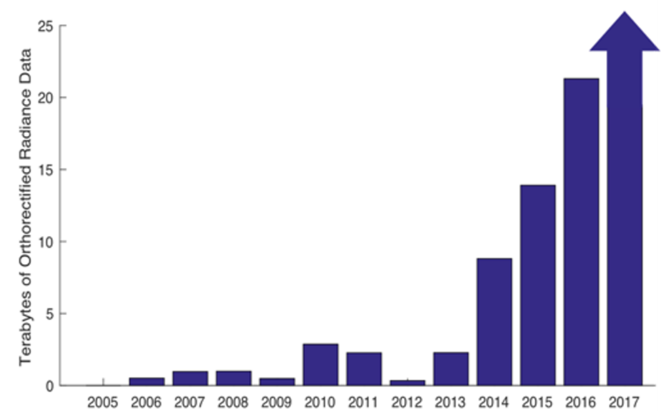
- “Chemical Fingerprint of the Atmosphere and Earth’s Surface” = dozens of information value-added products
- Backwards compatible with a 30-year record = huge existing user base
- Barriers to use: Big data – volumes and processing



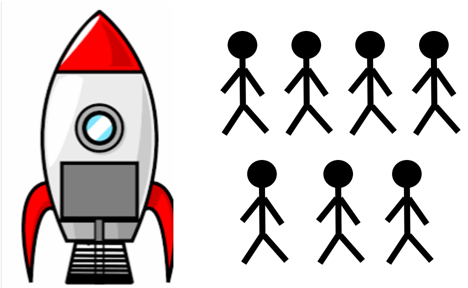
← Spectroscopy

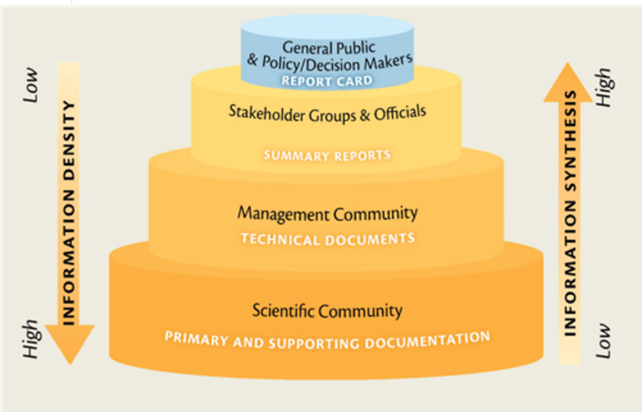
MultiSpectral →





SBG Data is Valuable to a Broad Range of Users





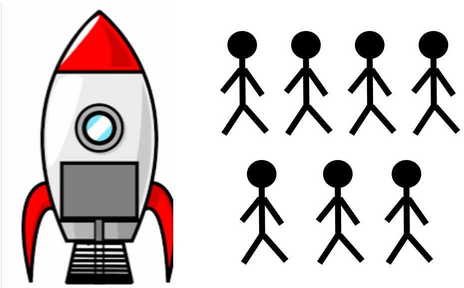
Types of Users

- Field scientist mostly interested in remote data at point locations
- Remote Sensing algorithm developer
- Technician providing data processing support
- Researcher using remote sensing data as a tool in statistical and modeling experiments
- Software developer generating tools and services of value
- Manager using data visualizations to inform decisions
- Educator/Student working on limited scope class projects

Disciplines

- Hydrological Cycles, Water Resources and Aquatic Ecosystems
- Weather and Air Quality
- Terrestrial Ecosystems and Natural Resource Management
- Climate variability and Change
- Earth Surface/ Geology

ImgSPEC focuses on a Subset of SBG users



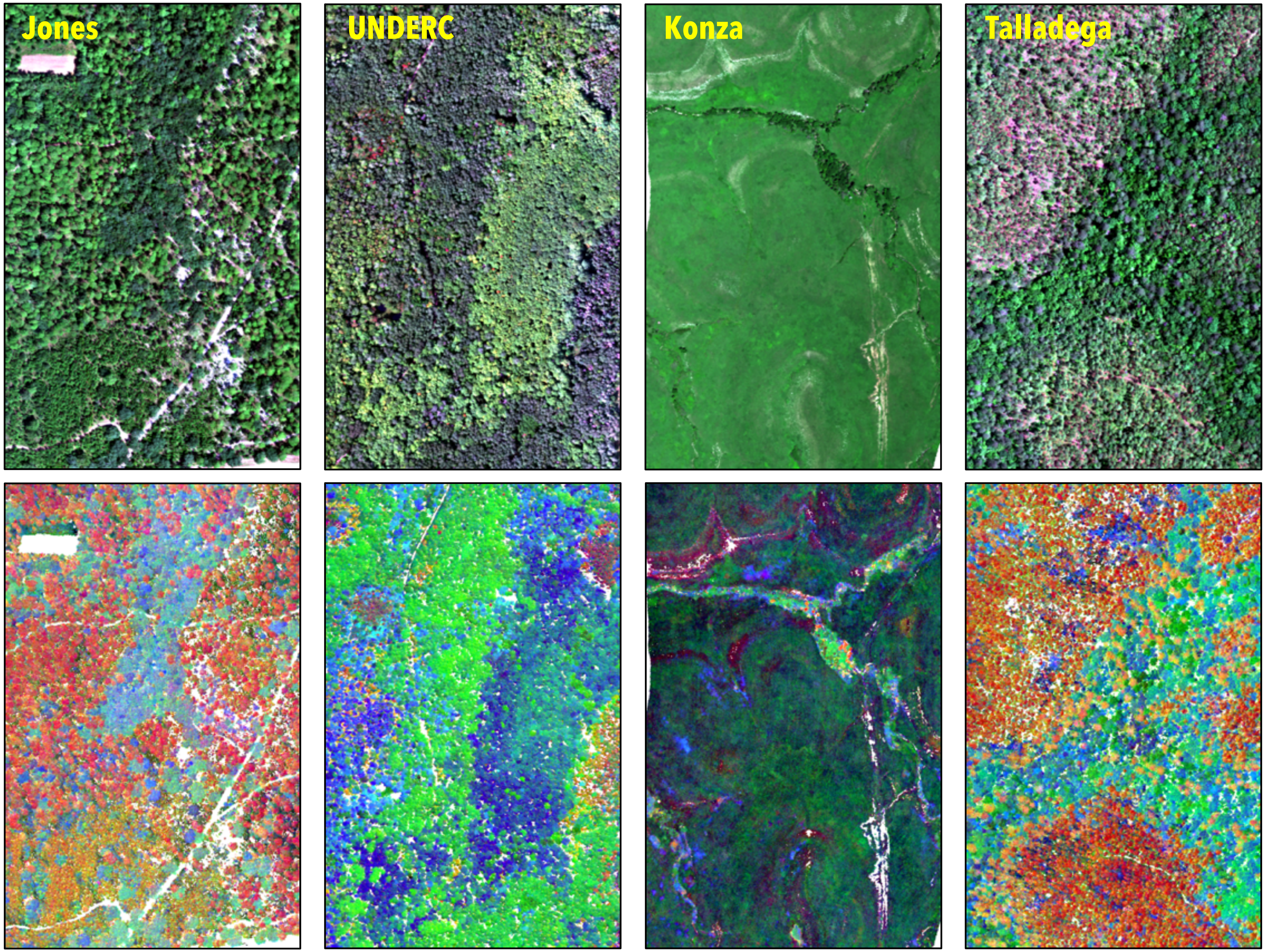
Types of Users

- Field scientist mostly interested in remote data at point locations
- Remote Sensing algorithm developer
- Technician providing data processing support
- Researcher using remote sensing data as a tool in statistical and modeling experiments
- Software developer generating tools and services of value
- Manager using data visualizations to inform decisions
- Educator/Student working on limited scope class projects

Disciplines

- Hydrological Cycles, Water Resources and Aquatic Ecosystems
- Weather and Air Quality
- Terrestrial Ecosystems and Natural Resource Management
- Climate variability and Change
- Earth Surface/ Geology

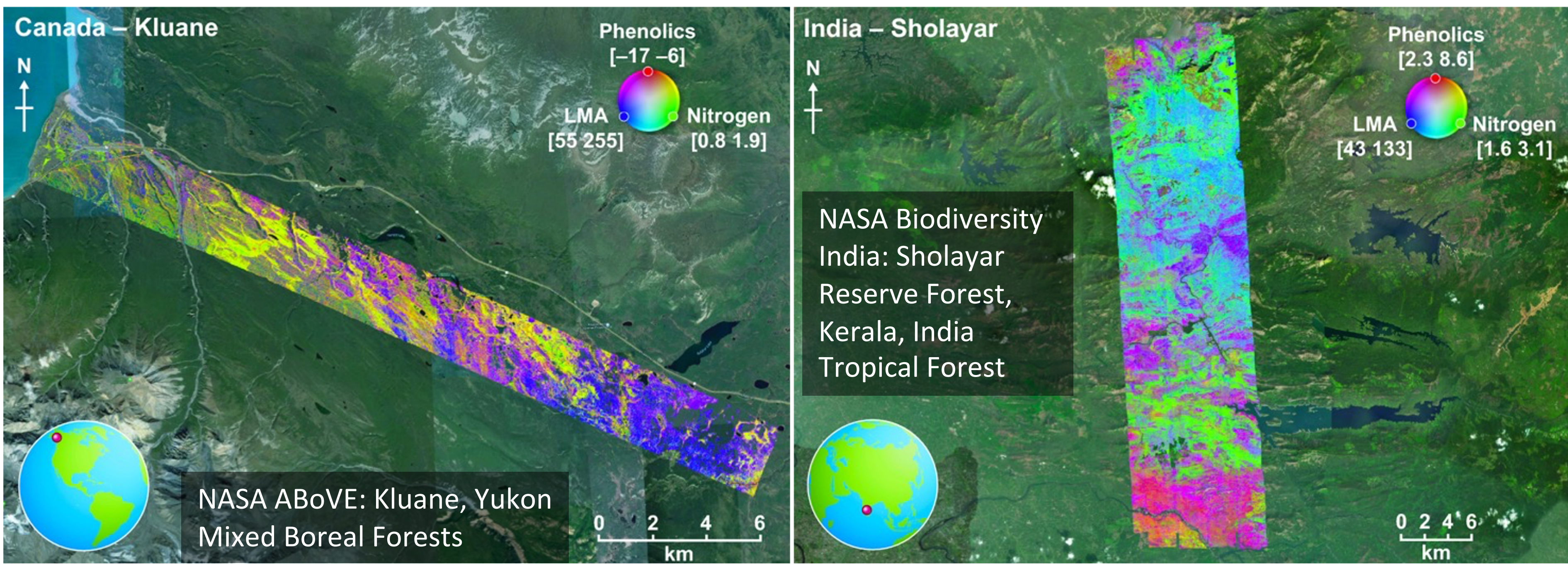
Use Case: Foliar Traits (a key SBG product)



R: LMA
G: Chl
B: Phenolics

ImgSPEC Objective

For the terrestrial ecology use case, demonstrate an end-to-end, on-demand, processing platform on the cloud for imaging spectroscopy Level 1 calibrated radiance data through Level 3+ information products



Schimel, Schneider et al. 2019, *New Phytologist*

Problem we want to solve: many plausible algorithms for same "product"


Remote Sensing of Environment 158 (2015) 15–27

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

ELSEVIER

Quantifying forest canopy traits: Imaging spectroscopy versus field survey

Gregory P. Asner

Department of Global Ecology

Research

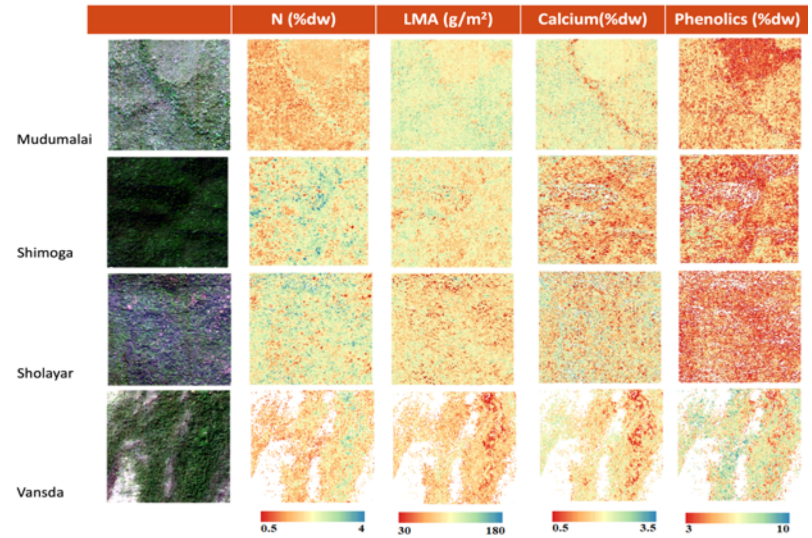
New Phytologist

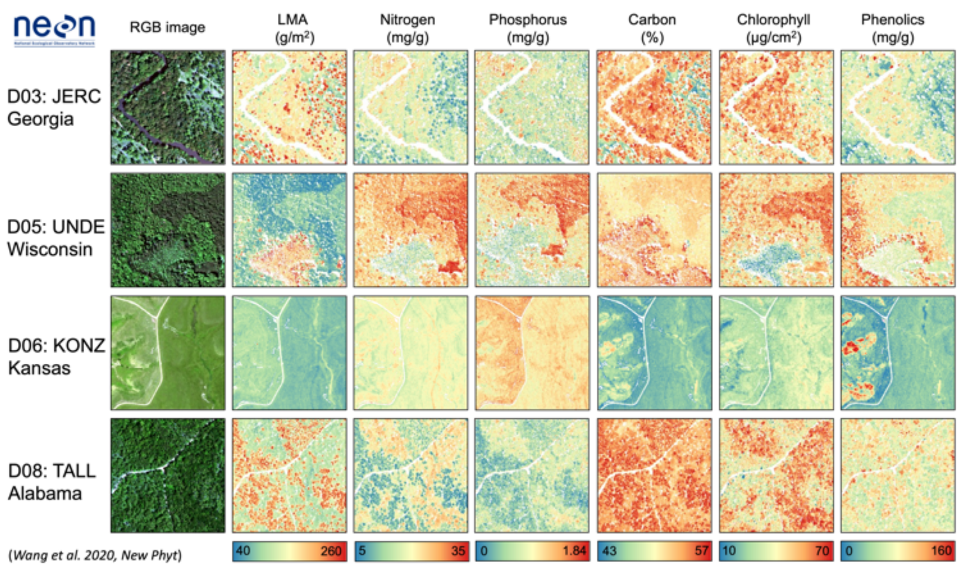
Foliar functional traits from imaging spectroscopy across biomes in eastern North America

Zhihui Wang[1], Adam Chlus[1], Ryan Geygan[1], Zhiwei Ye[1], Ting Zheng[1], Aditya Singh[2], John J. Couture[3], Jeannine Cavender-Bares[4], Eric L. Kruger[1] and Philip A. Townsend[1]

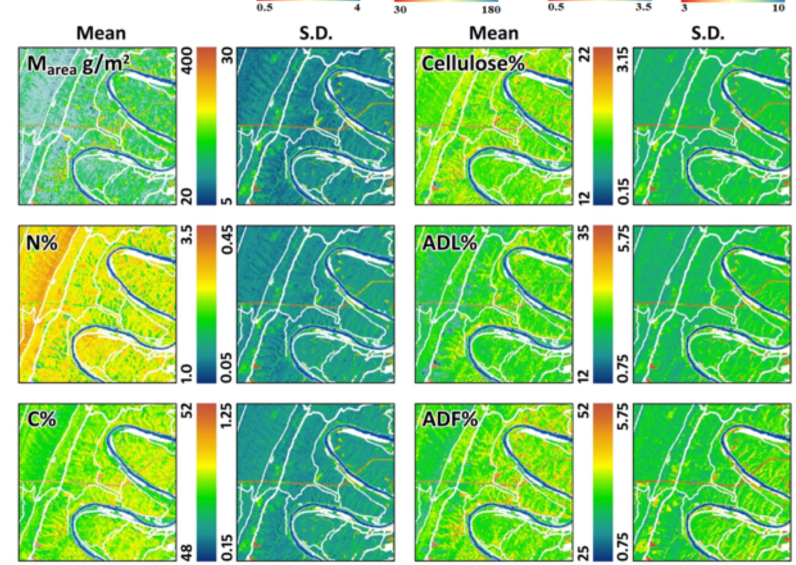
Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties

Aditya Singh,[1,3] Shawn P. Serbin,[1,4] Brenden E. McNeil,[2] Clayton C. Kingdon,[1] and Philip A. Townsend[1]

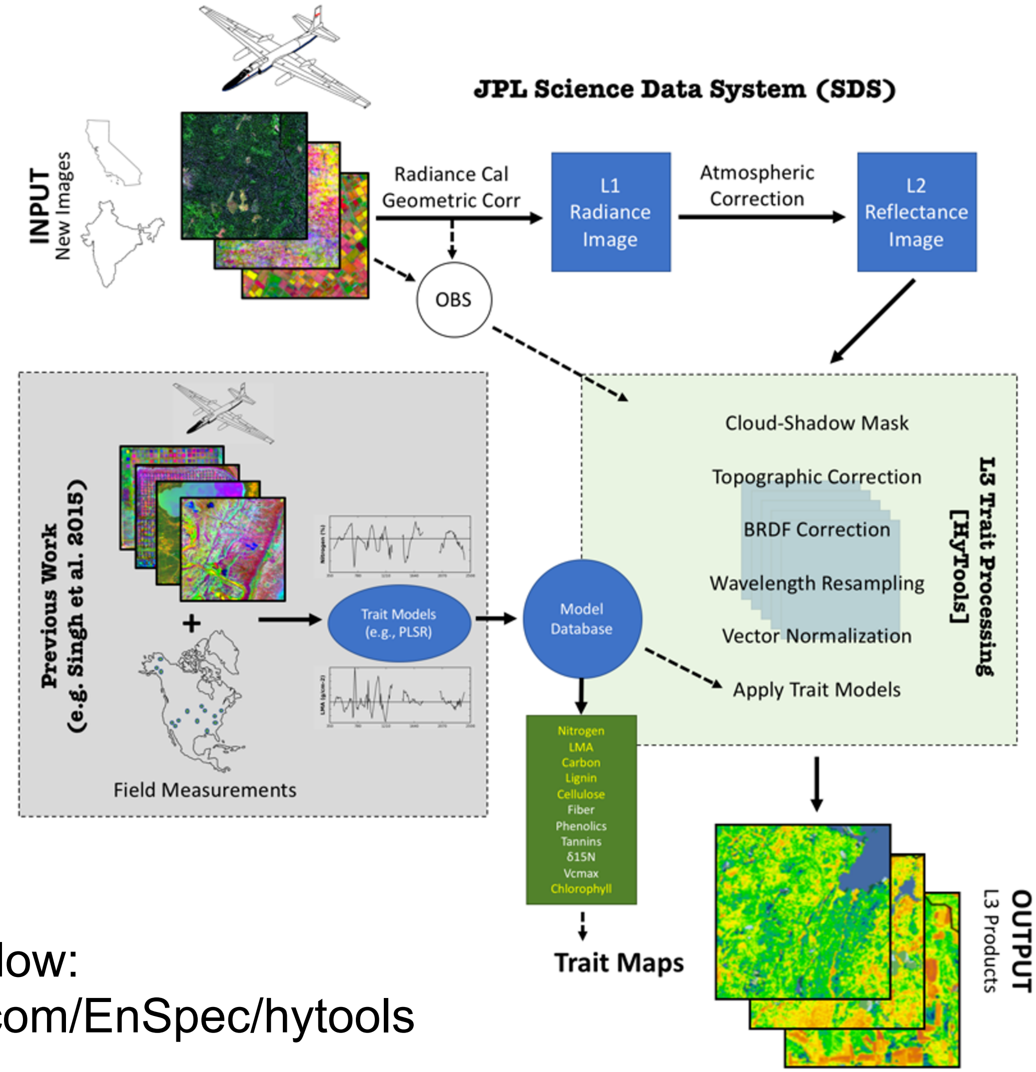




(Wang et al. 2020, New Phyt)

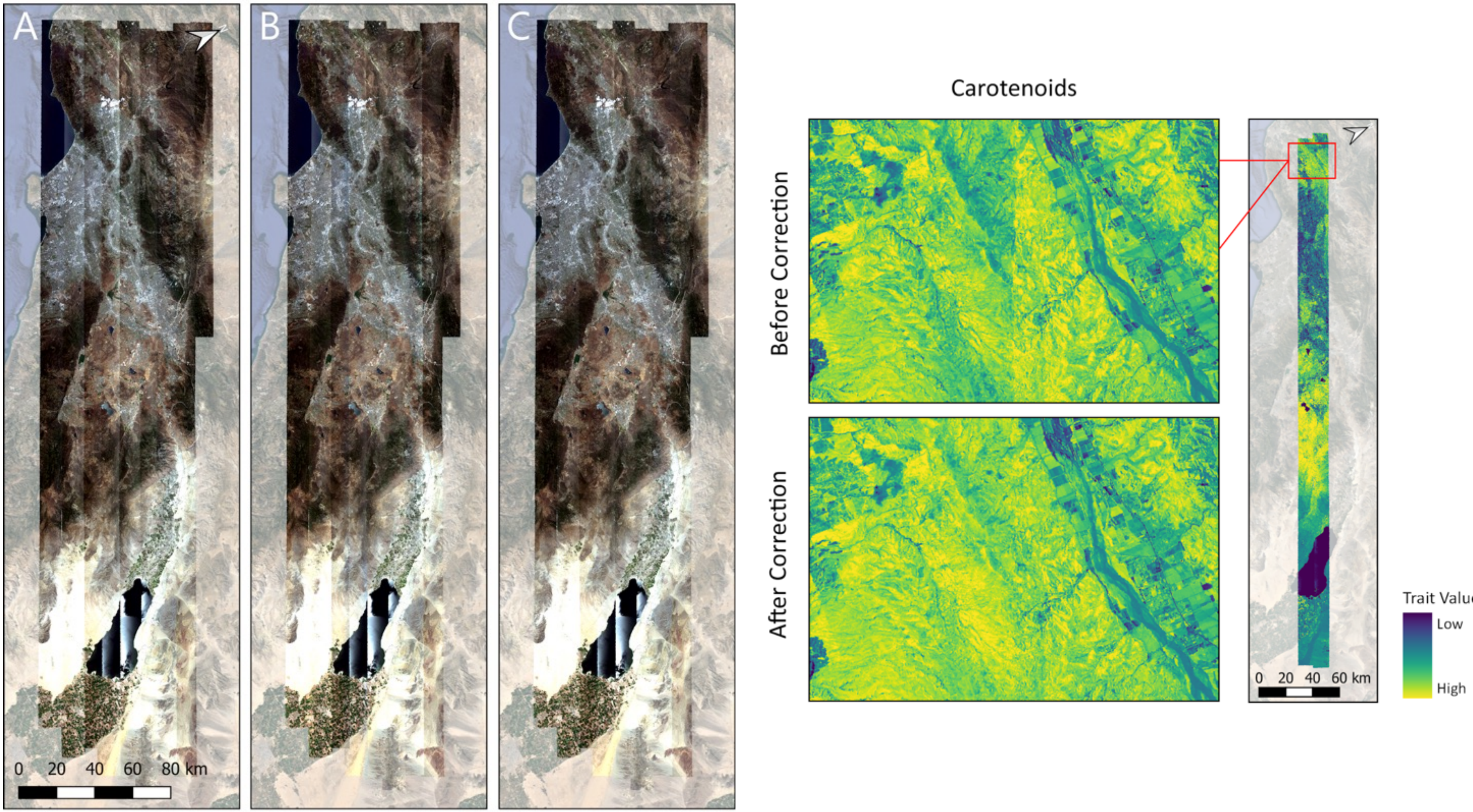


Problem we want to solve: multiple different processing workflows



HyTools workflow:
https://github.com/EnSpec/hytools

Intermediate processing (e.g. BRDF correction)



Queally et al. *in prep*

Context and Setting the Stage for Infusion

- The ImgSPEC premise is that data distribution from SBG will differ dramatically from current approaches; specifically, in relation to the number of potential products, algorithms, and pre-processing steps
  – Aligns with ongoing discussions with Kevin Murphy about a science data system as a service
  – Builds on Multi-mission Algorithm and Analysis Platform (MAAP)
  – Builds on another AISTs: 1) Smart, on-demand for multi-temporal SAR; 2) GJAM, 3) MOL
- Rather than locking users into a specific processing flow, ImgSPEC provides on-demand, customized processing workflows that support SBG Pathfinder SISTER:
  – Maintain provenance
  – Enable Reproducibility
  – Limit data download bottlenecks
  – Limit scope of development of all possible SBG products
  – Limit costs for reprocessing an entire data set when algorithms improve

Programmatic Objectives

- Develop an on-demand SDS for distributing new and existing L2+ imaging spectroscopy data from:
  - AVIRIS-classic and Next Generation, which currently distributes L1 and L2
  - NEON
  - SBG, a “designated” spaceborne global imaging spectrometer (ESAS, 2017) with >100 different potential algorithms for Level 1+ data products

- Expand use of existing imaging spectroscopy data by non-remote sensing experts:
  - Functional diversity and vegetation evolutionary lineage (collaborator *Jeannine Cavender-Bares*)
  - Predictive mapping of animal distributions (collaborator *Walter Jetz:* MOL)
  - Quantify aspects of ecosystem function (collaborators *Jennifer Swenson/Jim Clark:* GJAM)

- Relevant to multiple programs in the NASA Earth Science Division:
  - Direct: Terrestrial Ecology (Hank Margolis, ABoVE), Biological Diversity (Woody Turner), Ecological Forecasting (Woody Turner) [SBG Precursor], Land Cover/Land Use Change (Garik Gutman)
  - Indirect: Ocean Biology and Biogeochemistry (Laura Lorenzoni) and Water Resources (Brad Doorn)

- Leverage previously funded NASA tech: EcoSIS, EcoSML, HySDS, and SDAP

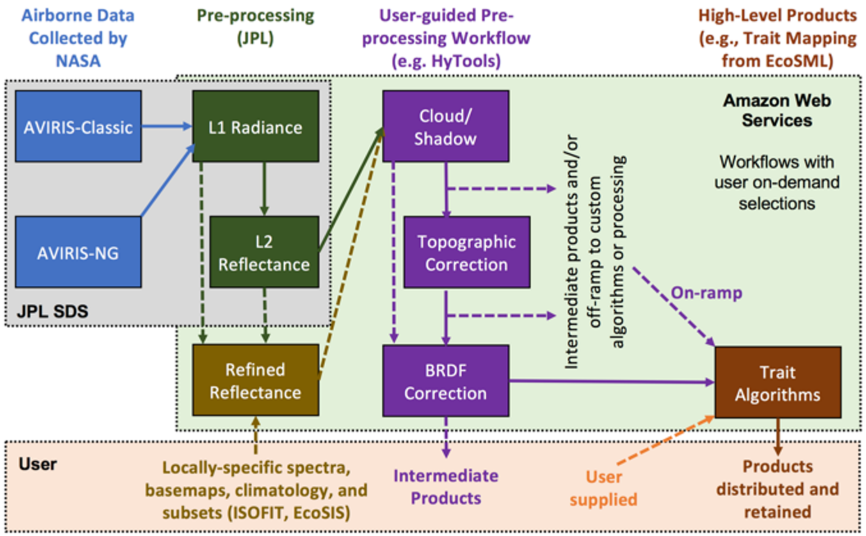
Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

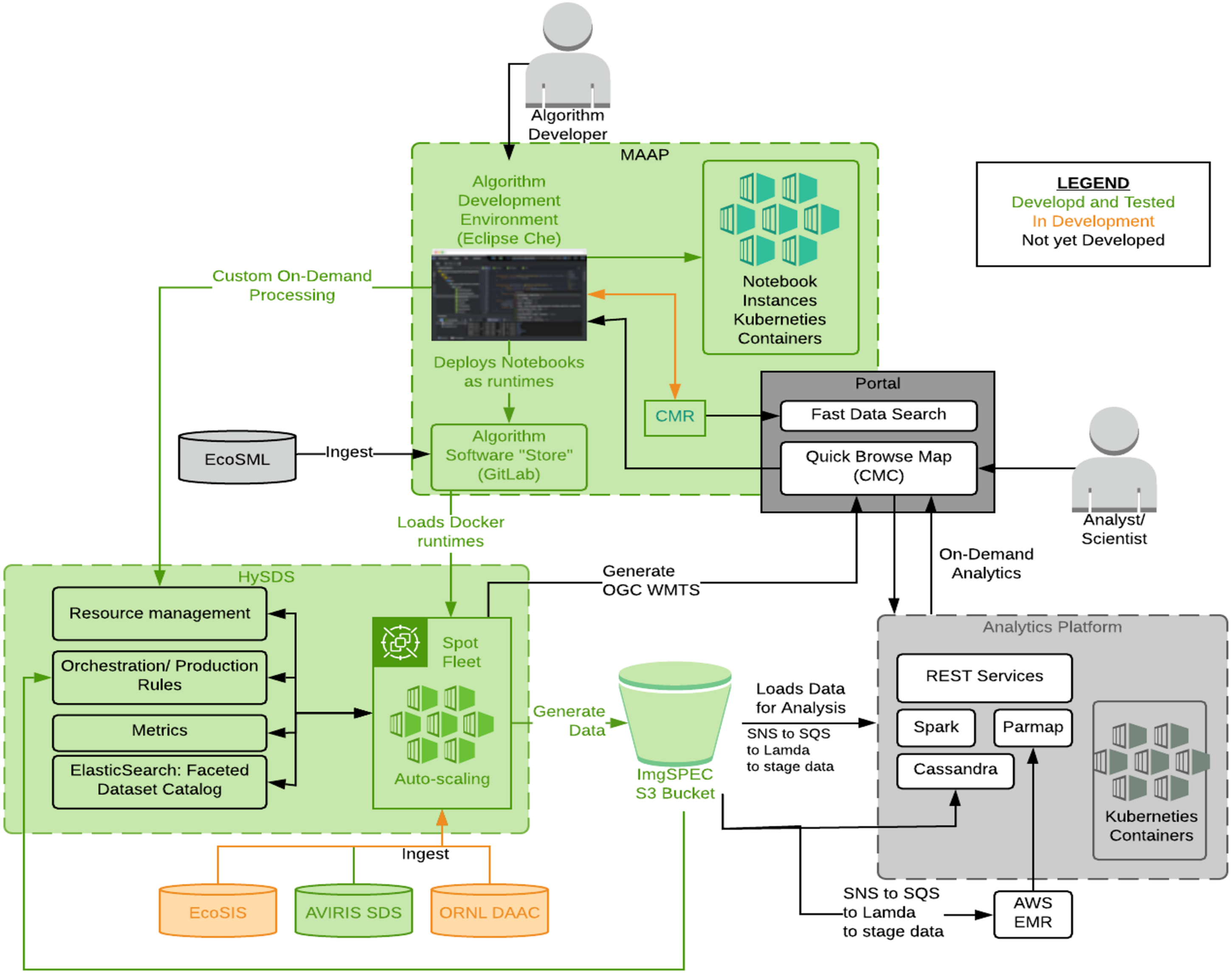
Technical Advancements

Prototype on-demand AVIRIS SDS that:

1. Automatically ingests data from the AVIRIS SDS and ORNL DAAC
2. Processes user-defined, on-demand Level 1 to Level 3 data
3. Enables three levels of user interaction:
   - prescribed start-to-finish workflow with existing algorithms
   - prescribed workflow with user parameterization or algorithm selection for processing at different stages
   - workflow that allows user off-ramps and on-ramps to the processing
4. Includes metadata to maintain provenance and enable reproducibility
5. Provides a GUI for visualizing the data and produce on-the-fly data analytics and cloud optimized GeoTIFFs
6. Enables download of a CSV containing data used in the analytic plots
7. Open-source and archive ImgSPEC software



Technical Advancements: ImgSPEC design leverages component technologies to create an on-demand, cloud-based science data system



Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

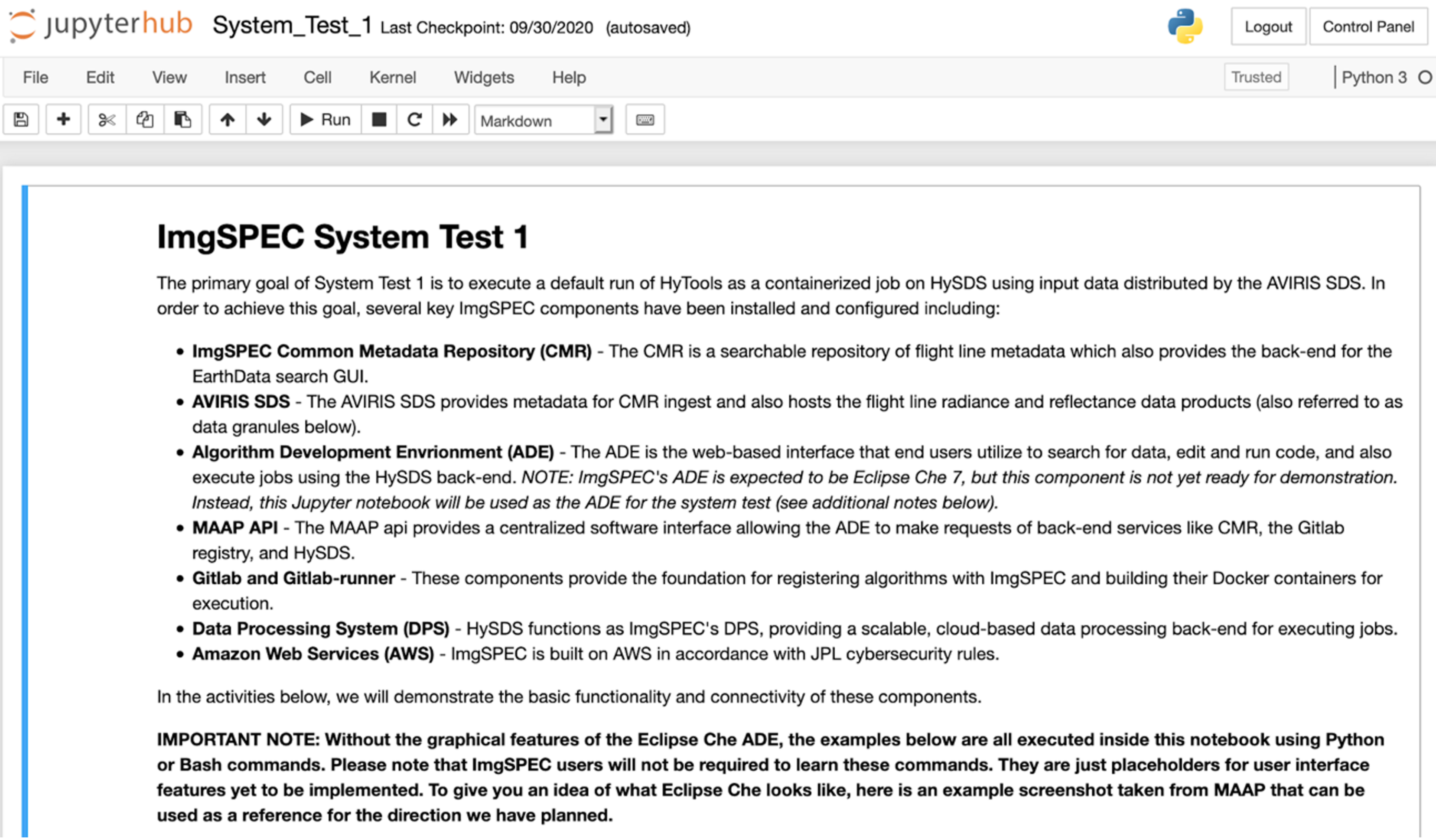
6-month Update Overview

- System Tests 1 & 2
- Documentation:
  - Metrics for Return on Investment
  - Interface Control Document
- Data formats and metadata standards for imaging spectroscopy and on-demand data product generation
- Continued partnerships with collaborator AISTs, JPL Unity and SBG Pathfinder (SISTER)

System Test 1

- **ImgSPEC Common Metadata Repository (CMR)**
- **AVIRIS SDS Interface**
- **Algorithm Development Envrionment (ADE)** - The ADE is the web-based interface that end users utilize to search for data, edit and run code, and also execute jobs using the HySDS back-end
  - *NOTE: Eclipse Che 7 not yet ready implemented; instead, this Jupyter notebook used for Test 1*
- **MAAP API** - centralized software interface allowing the ADE to make requests of back-end services like CMR, the Gitlab registry, and HySDS
- **Gitlab and Gitlab-runner** - Foundation for registering algorithms with ImgSPEC and building their Docker containers for execution
- **Data Processing System (DPS)** - HySDS functions as ImgSPEC's DPS, providing a scalable, cloud-based data processing back-end for executing jobs
- **Amazon Web Services (AWS)** - ImgSPEC is built on AWS in accordance with JPL cybersecurity rules

Jupyter Notebook Deployment



ImgSPEC System Test 1

The primary goal of System Test 1 is to execute a default run of HyTools as a containerized job on HySDS using input data distributed by the AVIRIS SDS. In order to achieve this goal, several key ImgSPEC components have been installed and configured including:

- **ImgSPEC Common Metadata Repository (CMR)** - The CMR is a searchable repository of flight line metadata which also provides the back-end for the EarthData search GUI.
- **AVIRIS SDS** - The AVIRIS SDS provides metadata for CMR ingest and also hosts the flight line radiance and reflectance data products (also referred to as data granules below).
- **Algorithm Development Envrionment (ADE)** - The ADE is the web-based interface that end users utilize to search for data, edit and run code, and also execute jobs using the HySDS back-end. *NOTE: ImgSPEC's ADE is expected to be Eclipse Che 7, but this component is not yet ready for demonstration. Instead, this Jupyter notebook will be used as the ADE for the system test (see additional notes below).*
- **MAAP API** - The MAAP api provides a centralized software interface allowing the ADE to make requests of back-end services like CMR, the Gitlab registry, and HySDS.
- **Gitlab and Gitlab-runner** - These components provide the foundation for registering algorithms with ImgSPEC and building their Docker containers for execution.
- **Data Processing System (DPS)** - HySDS functions as ImgSPEC's DPS, providing a scalable, cloud-based data processing back-end for executing jobs.
- **Amazon Web Services (AWS)** - ImgSPEC is built on AWS in accordance with JPL cybersecurity rules.

In the activities below, we will demonstrate the basic functionality and connectivity of these components.

IMPORTANT NOTE: Without the graphical features of the Eclipse Che ADE, the examples below are all executed inside this notebook using Python or Bash commands. Please note that ImgSPEC users will not be required to learn these commands. They are just placeholders for user interface features yet to be implemented. To give you an idea of what Eclipse Che looks like, here is an example screenshot taken from MAAP that can be used as a reference for the direction we have planned.

System Test 2

- Introduce Algorithm Development Environment (ADE) - main user interface to develop algorithms, locate data, and execute jobs through many workspace options (Jupyter, RStudio, etc)
- Use Earthdata search client
- Demonstrate ISOFIT fully-open source atmospheric correction registration with ImgSPEC in ADE
- Run a ISOFIT job through ADE to backend
- Demonstrate final publication of data products to Earth data search

Metrics for Return on Investment

- Reduced download times (Req 3)
  - Download time for flightline/flight box from FTP vs. Download AWS
  - Download time for flightline/flight box vs. download CSV at relevant points from AWS

- Easy provenance for reproducibility (Req 7)
  - Reproducible L2+ product from given metadata

- Scalable work environments (Req 4-6)
  - Processing time for variable size jobs as an AWS job vs. AVIRIS SDS job

Imaging Spectroscopy
Data Formats and Metadata

- Coordinating with MEET-SBG and SISTER
- Defining UMM-C and UMM-G science key words
- Data Formats:
  - Assume ENVI format coming in
  - Work in COG
  - Still TBD: Export to nc4 or ENVI format
- Metadata: What is needed by scientists to reproduce a granule?
  - Link to code that generated it including all subsequent data product levels: “PGE + version”
  - Coefficients (parameters for a model):
  - Ancillary data -- “input granules”

Plan Forward

Aug 2020 System Test 3: Interface with EcoSIS

Start Date	Due Date	What
9/1/2020	1/30/2021	Read ECOSIS API into ImgSPEC ADE
1/15/2021	2/7/2021	Setup iframe in ADE of EcoSIS
2/7/2021	2/14/2021	Search for EcoSIS spectra through iframe in ADE
2/14/2021	2/21/2021	Use EcoSIS spectra in ADE in ISOFIT
1/15/2021	2/21/2021	Define write spectra to EcoSIS interface
1/15/2021	2/21/2021	ORNL Interface: ORNL data ingest to ImgSPEC search
1/15/2021	2/14/2021	Implement ImgSPEC ADE login through UAT-URS
1/15/2021	2/21/2021	Set-up new HySDS system to enable auto-scaling
1/1/2021	2/21/2021	System Test 3: ISOFIT with EcoSIS spectra
2/21/2021	2/28/2021	System Test 3: User Acceptance Testing

Summary

- COVID complicated staffing
  - As of October fully staffed at JPL and U Wisconsin
  - JPL Co-I on extend leave returned
  - JPL Product Delivery Manager moving to CU Boulder as Co-I; new PDM
  - Ramping up FTE to mitigate risks and catch-up on spending
- System Test 1 Completed
  - Descoped to Jupyter Notebook with Eclipse Che Deployment deferred to Test 2
  - User Acceptance Testing
- System Test 2 Completed
  - Tested Algorithm Development Environment and HySDS interface with fully-open source ISOFIT
  - Demonstrated CMR faceted search and interfaces
- Continued collaborations with fellow AIST projects, SBG Pathfinder (SISTER), JPL Unity, and MAAP

Presentation Contents

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

Publications/ Conferences

- Publication in draft on the user-centered design approach with ImgSPEC as the case study (Aim: Research Technology Management)
- Made the ESTO news story about coordinated Biodiversity AIST projects
- Attended ESTF 2020 virtually
- Keynote address planned at Apache Con in September by Co-I Huang
- Invited speaker at AIST session for AGU
- Spoke at AGU in virtual e-lightening

Acronyms

List of Acronyms

AGU	American Geophysical Union
API	Application Programming Interface
AVIRIS	Airborne Visible Infrared Imaging Spectrometer
BRDF	Bi-directional Reflectance Function
CMC	Common Mapping Client
collab	collaborator
CSV	Comma-Separated Values
EcoSIS	Ecological Spectral Information System
EcoSML	Ecological Spectral Model Library
ESAS	Earth Science and Applications from Space Decadal Survey
GeoSPEC	Geospatial Spectroscopy Processing Environment on the Cloud
HySDS	Hybrid Science Data System
ISOFIT	Imaging Spectrometer Optimal Fitting
NEON	National Ecological Observation Network
ORNL DAAC	Oak Ridge National Laboratory Data Active Archive Center
SBG	Surface Biology and Geology
SDAP	Science Data Analytics Platform
SDS	Science Data System

AMP: An Automated Metadata Pipeline

Beth Huffer (PI, Lingua Logica)
Damian Gessler (Co-I, Lingua Logica)
Simon Handley (Co-I, Lingua Logica)
Kala Babcock Gessler (Co-I, Lingua Logica)
Ken Bagstad (Co-I, USGS)
Ferdinando Villa (Co-I, Basque Center for Climate Change)
Jennifer Wei (Collaborator, NASA GES DISC)

AIST-18-0042 Annual Technical Review
02-05-2021

ESTO Earth Science Technology Office

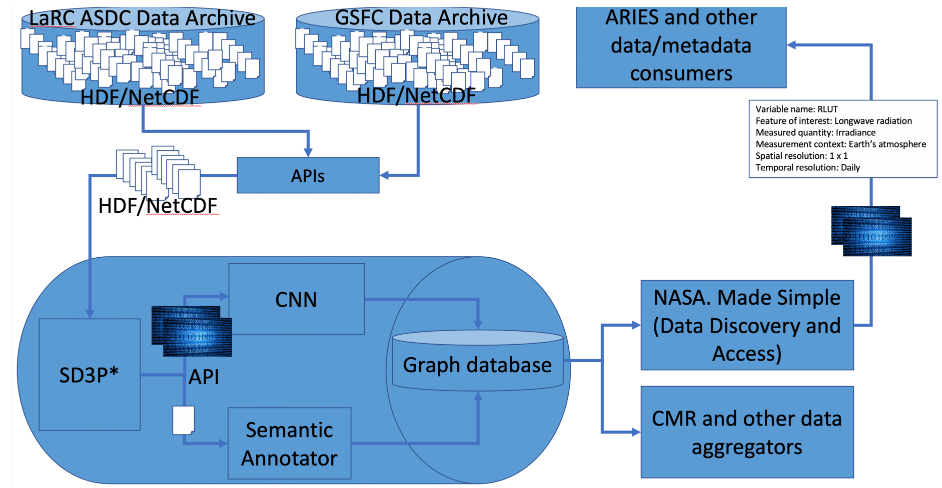
AMP: An Automated Metadata Pipeline

PI: Beth Huffer, Lingua Logica

Objectives

Develop an automated metadata pipeline (AMP) to create accurate, robust metadata that will facilitate data discovery and interoperability across Earth science data sources

- Use Machine Learning (ML) techniques to classify datasets ontologically and find related data
- Make NASA Earth science data easily discoverable and accessible
- Integrate AMP with Goddard Earth Science Data and Information Services Center (GES DISC) data distribution services
- Test usability of metadata and data services with the ARtificial Intelligence for Ecosystem Services (ARIES) modeling platform to demonstrate the value of semantics in improving usability



Automated Metadata Pipeline (AMP)

*Science Data Pre-Processing Package

Approach

Manually curate sample datasets with semantic annotations that uniquely characterize the data and assign the manually curated datasets to ontological classes in the AMP ontology

Using manually curated datasets, train Convolutional Neural Network (CNN) models to detect patterns in data to classify new data sets

Auto-generate semantically annotated metadata records for datasets classified by AMP for use by downstream systems such as ARIES and Common Metadata Repository (CMR)

Create interfaces between AMP and ARIES and GES DISC data services to enable deployment of AMP within ARIES and GES DISC

Co-Is/Partners: Co-Is/Partners: D. Gessler, S. Handley, C. Babcock-Gessler (Lingua Logica); K. Bagstad, USGS; F. Villa, Basque Center for Climate Change; J. Wei, GES DISC

Key Milestones

• AMP High Level System Design	05/20
• Decision tree for semantic parser/subsetter	08/20
• Finalize ontology design	09/20
• Demonstrate pre-alpha pipeline proof of concept	02/21
• Complete CNN training	02/21
• Finish end-to-end pipeline proof of concept	11/21

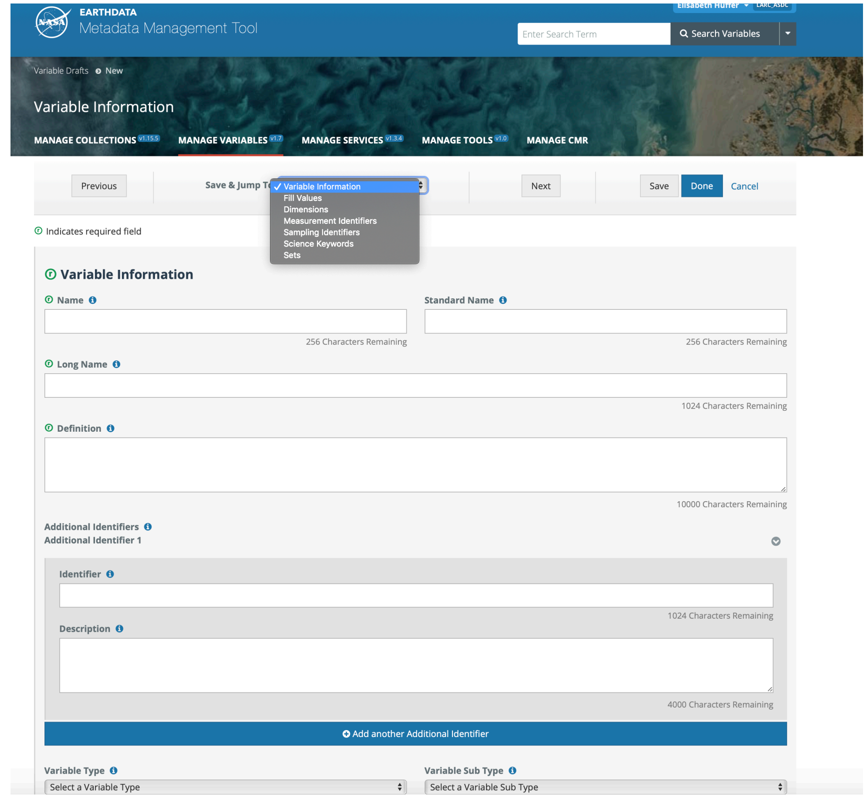
$TRL_{in} = 3$ $TRL_{current} = 4$

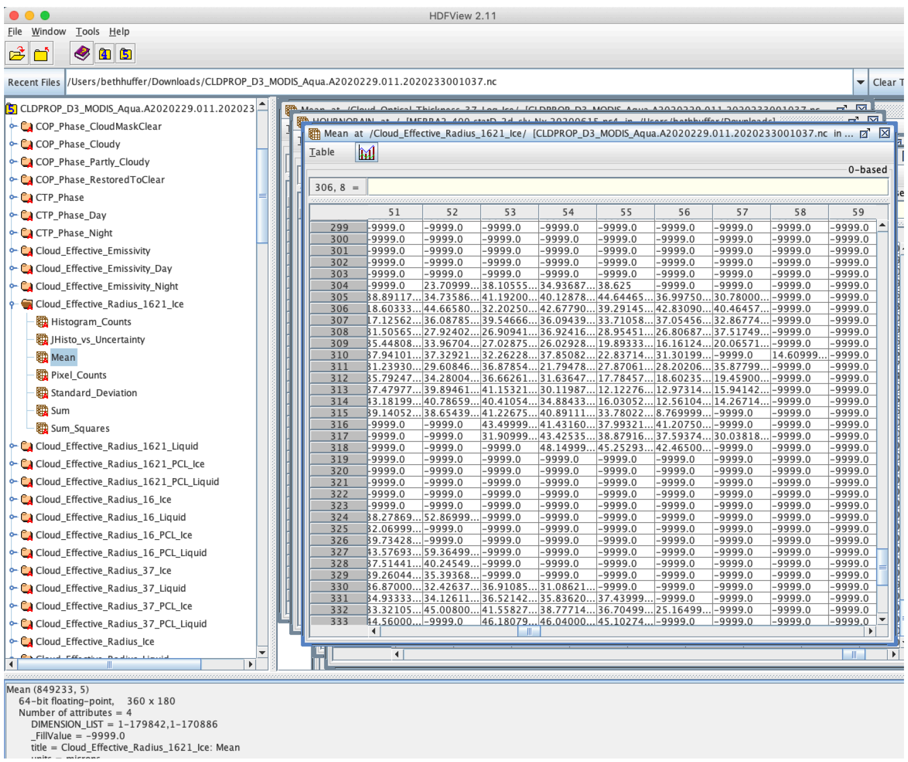
Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background and Objectives

- The primary goal of the AMP project is to make NASA Earth science data **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable
- FAIR data needs good metadata at the individual dataset level

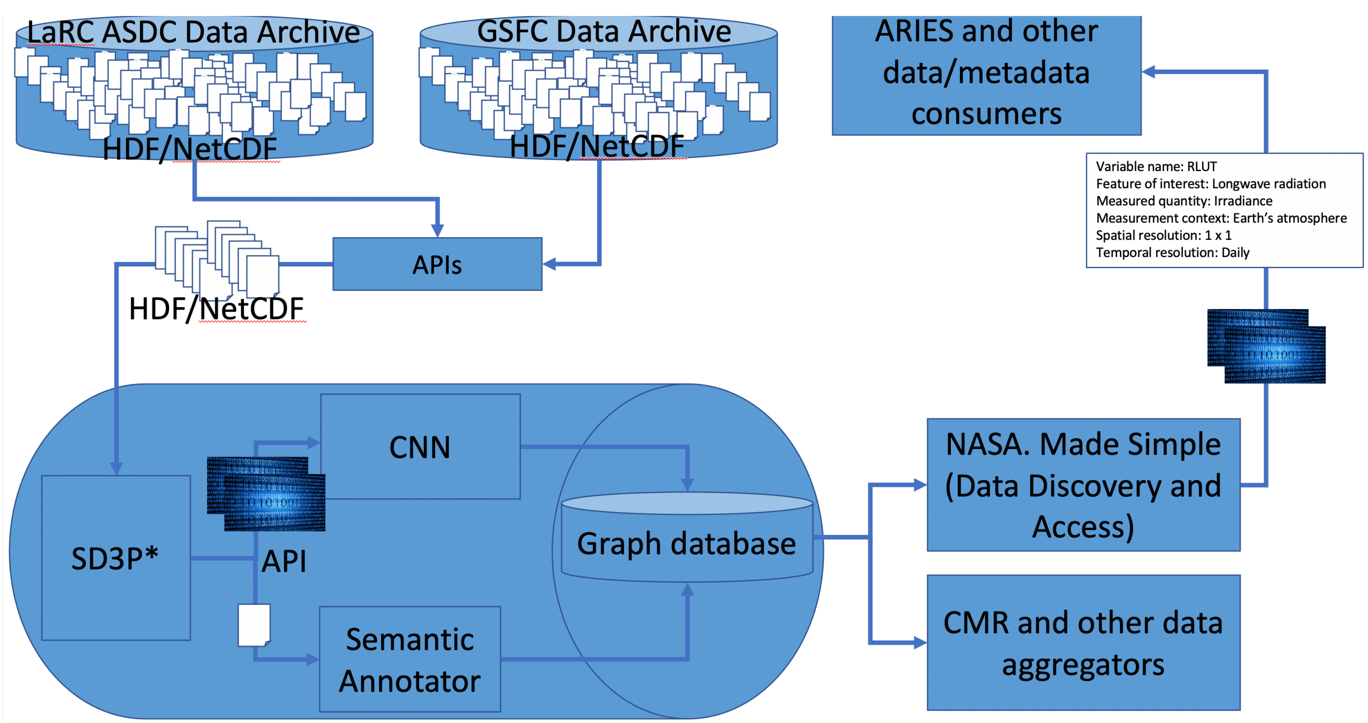




- NASA Earth science data archives contain 1000's of data collections (7500+ available at https://search.earthdata.nasa.gov
- Each collection contains multiple datasets
- Creating metadata records for each of these datasets is a daunting task; if it must be done manually, it probably won't get done

Background and Objectives

- AMP is an automated metadata production pipeline that combines ML and Semantics to produce the metadata needed to make data FAIR
- The AMP system has 3 main components:
  - An ML pipeline (CNN)
  - A Semantic Annotator + RDF Triple Store
  - The SD3P (Science Data Pre-Processing Package)



Automated Metadata Pipeline (AMP)

*Science Data Pre-Processing Package

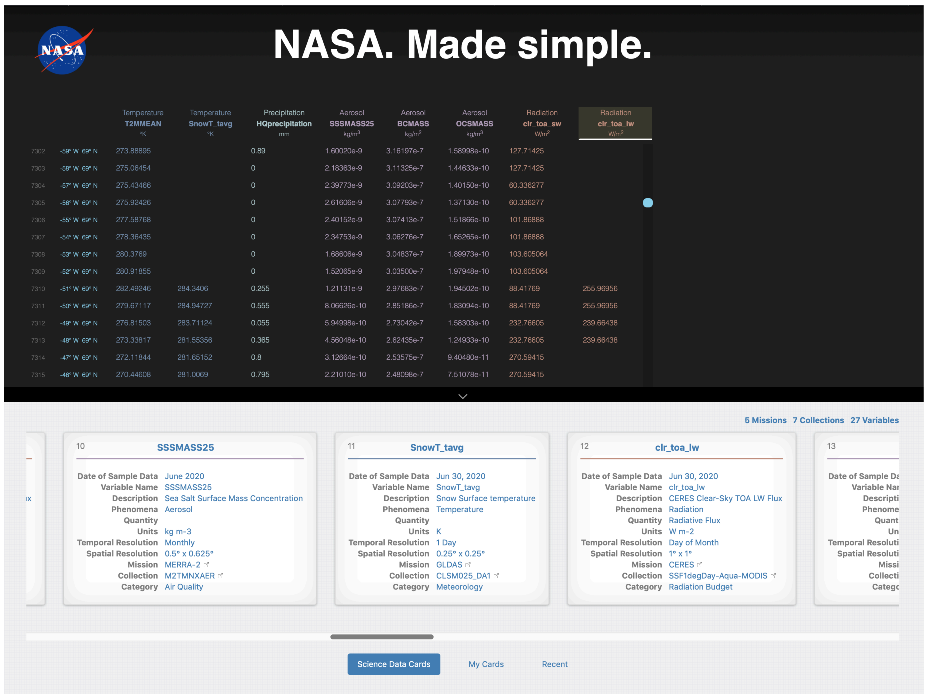
- The ML Pipeline and the Semantic Annotator together generate metadata records that are semantically and syntactically consistent, and provide the information users need to work with the data
- The SD3P supports data access and interoperability by parsing and subsetting HDF files, providing on-demand access to analysis-ready datasets

Presentation Contents

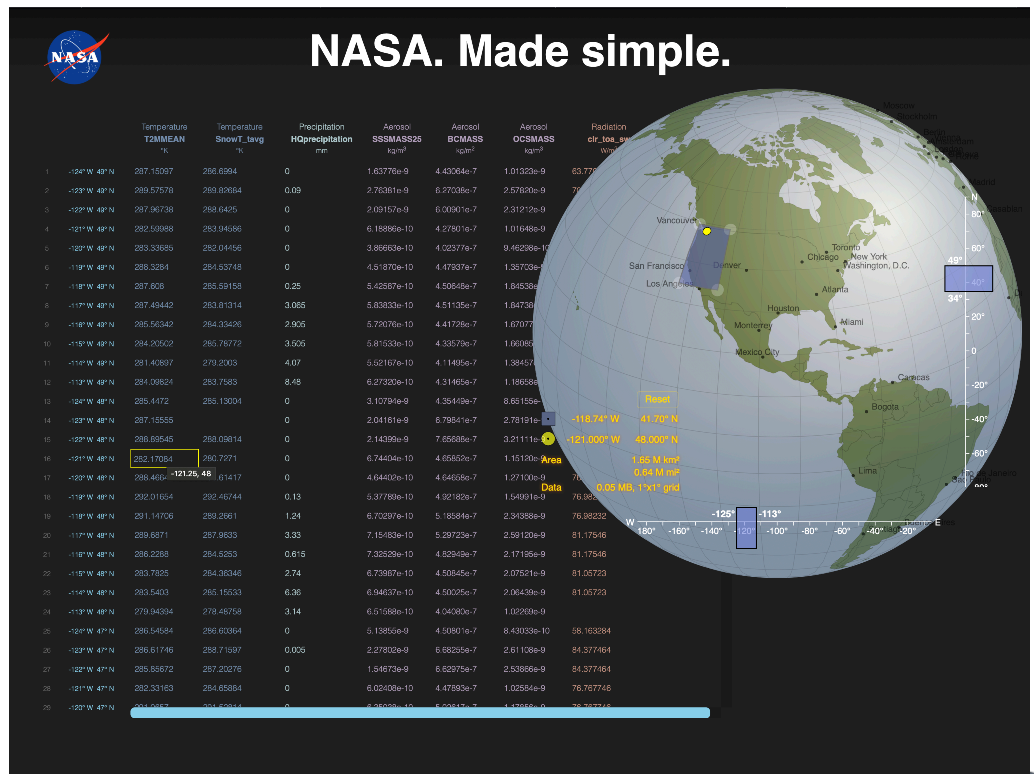
- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Background and Objectives

- Our pre-alpha version of the **NASA. Made Simple** portal demonstrates this for a broad sampling of NASA air quality, hydrology, meteorology, and radiation budget datasets



LIVE DEMO

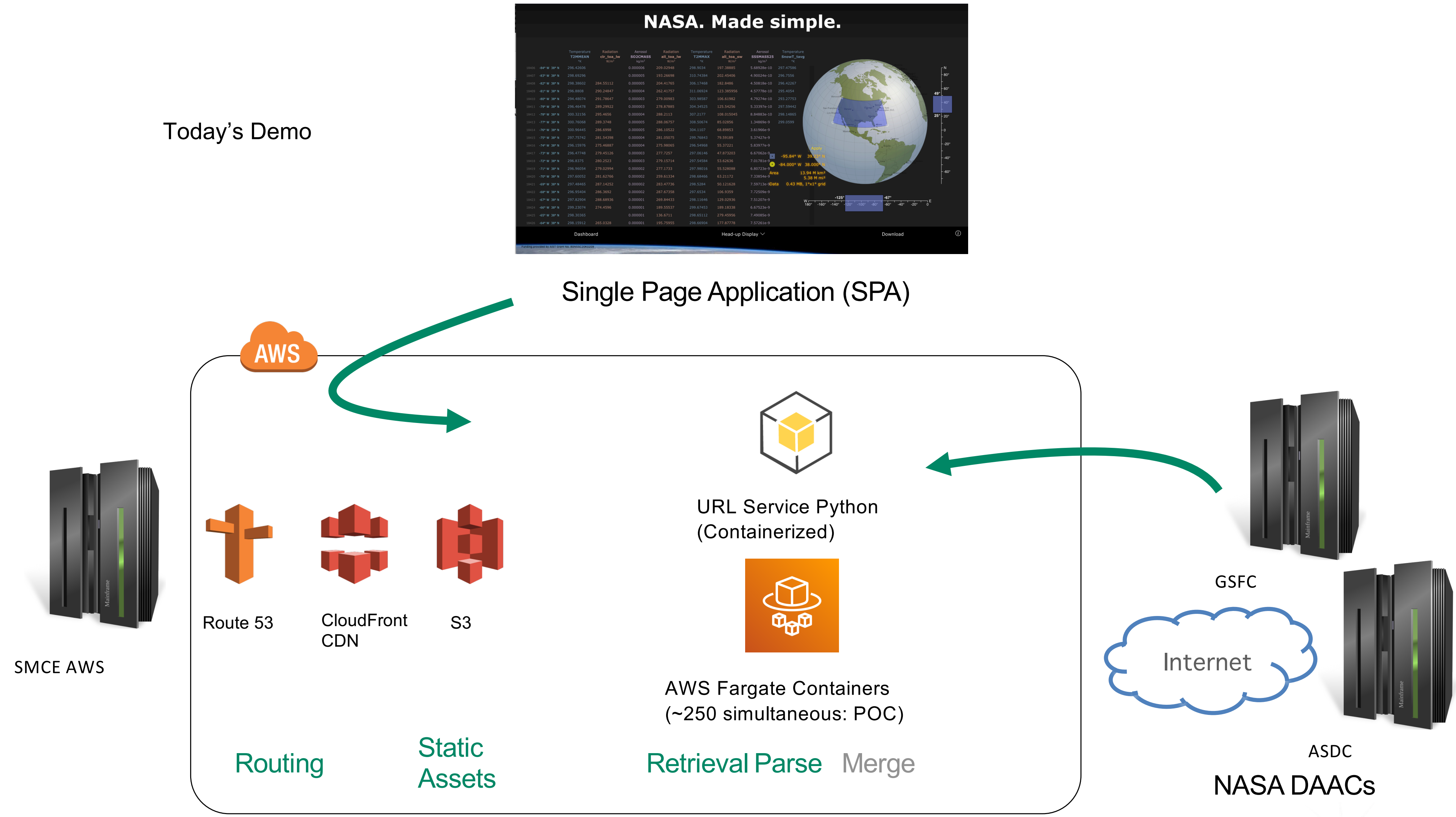


LIVE DEMO

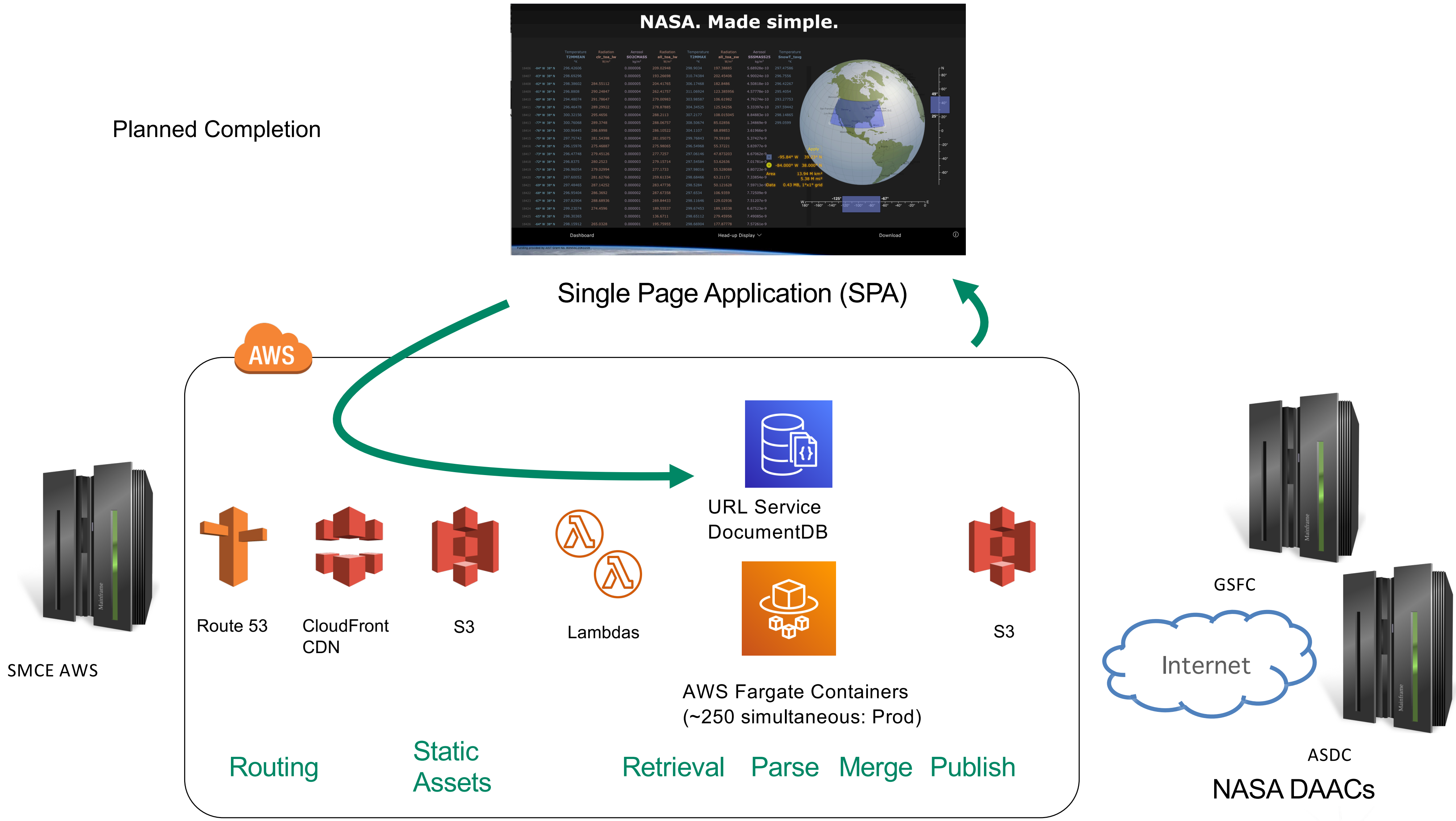
- HDF5 or HDF4 → HDF5
- Metadata: Destructuring of packed strings into structured data (JSON)
- Data: Filter missing and fill values; apply scale and offset transformations; attach units
- Geospatial: Projection extraction of measurements onto longitudinal and latitudinal coordinates
- Temporal: Lossless transformation to ISO-8601 standard ("time since January 1, 1948")
- Universal, multi-dimensional point-cloud:

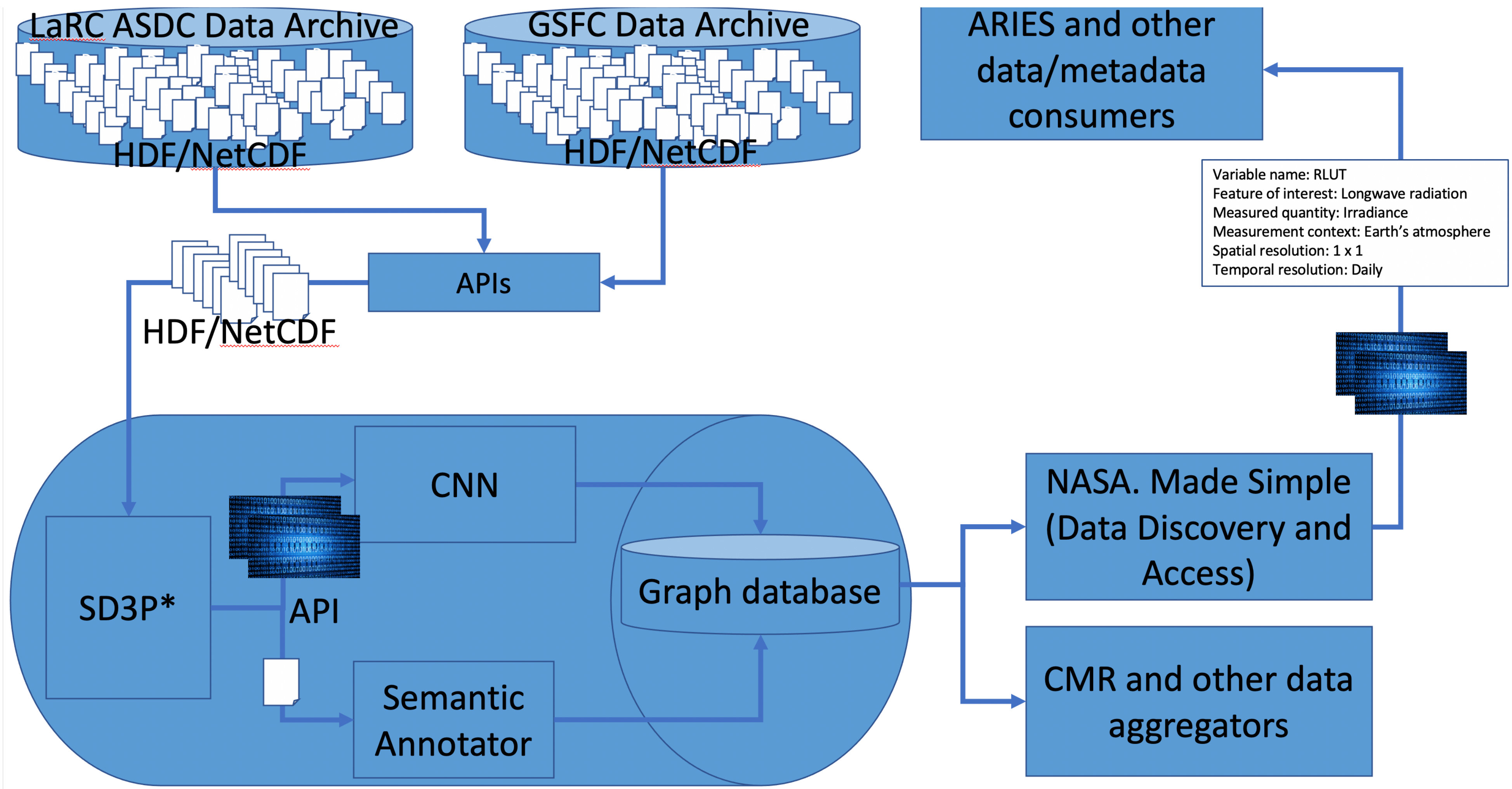
```
$ ./hdfp slice -n 10 /time /lat:[::1] /lon:[::1] /BCANGSTR /BCCMASS MERRA2_400.tavg1_2d_aer_Nx.20200630.nc4
/time   ISO-8601        /lat    /lon    /BCANGSTR       /BCCMASS
minutes since 2020-06-30 00:30:00       ISO-8601        degrees_north   degrees_east    1       kg m-2
24      24      361     576     4990464 4990464
0       2020-06-30T00:30:00     -89.0   -179.375        1.5657406       8.934125e-08
0       2020-06-30T00:30:00     -89.0   -178.125        1.5657406       8.934125e-08
0       2020-06-30T00:30:00     -89.0   -177.5  1.5657406       8.934125e-08
0       2020-06-30T00:30:00     -89.0   -176.25 1.5657406       8.934125e-08
0       2020-06-30T00:30:00     -89.0   -175.0  1.565176        8.9457664e-08
0       2020-06-30T00:30:00     -89.0   -174.375        1.5651379       8.9457664e-08
0       2020-06-30T00:30:00     -89.0   -173.125        1.5651379       8.9457664e-08
0       2020-06-30T00:30:00     -89.0   -172.5  1.5651379       8.9457664e-08
0       2020-06-30T00:30:00     -89.0   -171.25 1.5651379       8.9457664e-08
0       2020-06-30T00:30:00     -89.0   -170.0  1.5651379       8.9457664e-08
```

Cross-DAAC geospatial/temporal data integration across missions, collections, and variables


Today's Demo
NASA. Made simple.
Single Page Application (SPA)
AWS
Route 53
CloudFront
CDN
S3
URL Service Python
(Containerized)
AWS Fargate Containers
(~250 simultaneous: POC)
Routing
Static
Assets
Retrieval Parse
Merge
SMCE AWS
Internet
GSFC
ASDC
NASA DAACs

Planned Completion





Automated Metadata Pipeline (AMP)

*Science Data Pre-Processing Package

Technical and Science Advancements

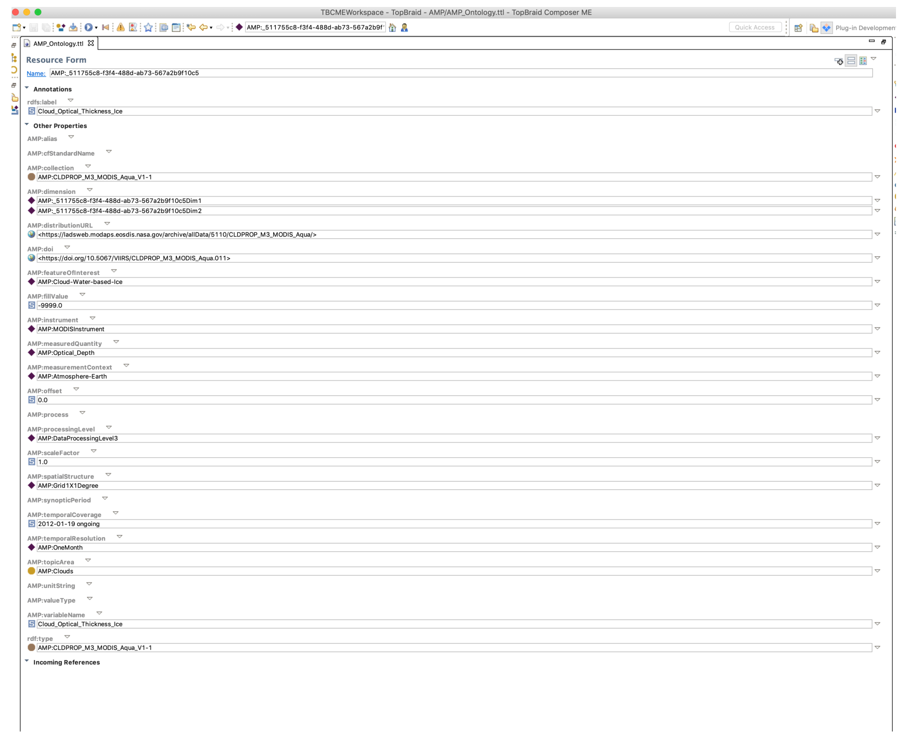
AMP metadata records support discovery of datasets that are highly relevant to an individual user's specific research needs because they include the information needed to use the data, including:

- The collection that the dataset belongs to
- The dimensions
- The urls for downloading the data, and the dataset landing page
- Fill values, offsets, and scale factors
- The spatial and temporal characteristics of the data (spatial/temporal coverage, spatial/temporal resolution)

• SD3P + Semantic Annotator

- The Earth science phenomenon represented by the data (what was measured, what quantity of it was measured, and what context was it measured in

ML Pipeline + Semantic Annotator



Earth Science Technology Office

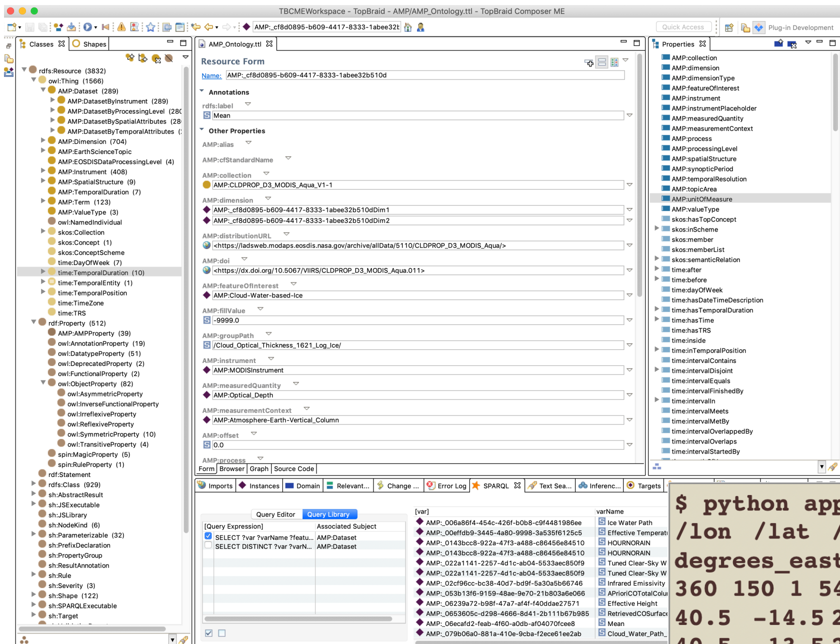
Technical and Science Advancements

- Completed and stabilized our system design
- Implemented prototypes of all 3 components

ML Pipeline

MERRA-2 air pollution datasets (part 1/2)

aer/BCSMASS	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.14)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SO2CMASS (0.02)	maer/SSSMASS25 (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
aer/DUCMASS25	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.14)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SO2CMASS (0.02)	maer/SSSMASS25 (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
aer/OCCMASS	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.14)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SO2CMASS (0.02)	maer/SSSMASS25 (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
aer/SO2CMASS	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.14)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SO2CMASS (0.02)	maer/SSSMASS25 (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
aer/SSSMASS25	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.14)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SO2CMASS (0.02)	maer/SSSMASS25 (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
maer/BCSMASS	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.15)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SSSMASS25 (0.02)	maer/SO2CMASS (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
maer/DUCMASS25	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.15)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SO2CMASS (0.02)	maer/SSSMASS25 (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
maer/OCCMASS	aer/BCSMASS (0.19)	aer/DUCMASS25 (0.18)	aer/SSSMASS25 (0.17)	aer/OCCMASS (0.15)	aer/SO2CMASS (0.14)	gpm-3/HQprecipitation (0.06)	maer/BCSMASS (0.02)	maer/SO2CMASS (0.02)	maer/SSSMASS25 (0.02)	maer/OCCMASS (0.02)	maer/DUCMASS25 (0.01)
maer/SO2C	aer/BCSMA	aer/DUCMASS25	aer/SSSMASS	aer/OCCMASS	aer/SO2CMASS	gpm-3/HQprecipitation	maer/BCSMASS	maer/SO2CMASS	maer/SSSMASS25	maer/OCCMASS	maer/DUCMASS25



SEMANTIC ANNOTATOR/
RDF Triple Store

```
$ python app.py get /lon:[40:50
/lon  /lat  /time /Albedo_inst  /A
degrees_east  degrees_north date
360 150 1 54000
40.5  -14.5 2014-01-01 20.882565
40.5  -13.5 2014-01-01 19.7368829
40.5  -12.5 2014-01-01 19.773048
40.5  -2.5  2014-01-01 15.350428
40.5  -1.5  2014-01-01 16.658976
40.5  -0.5  2014-01-01 17.662725
40.5  0.5   2014-01-01 17.940388
40.5  1.5   2014-01-01 16.766033
40.5  2.5   2014-01-01 15.799581
40.5  3.5   2014-01-01 14.354543
```

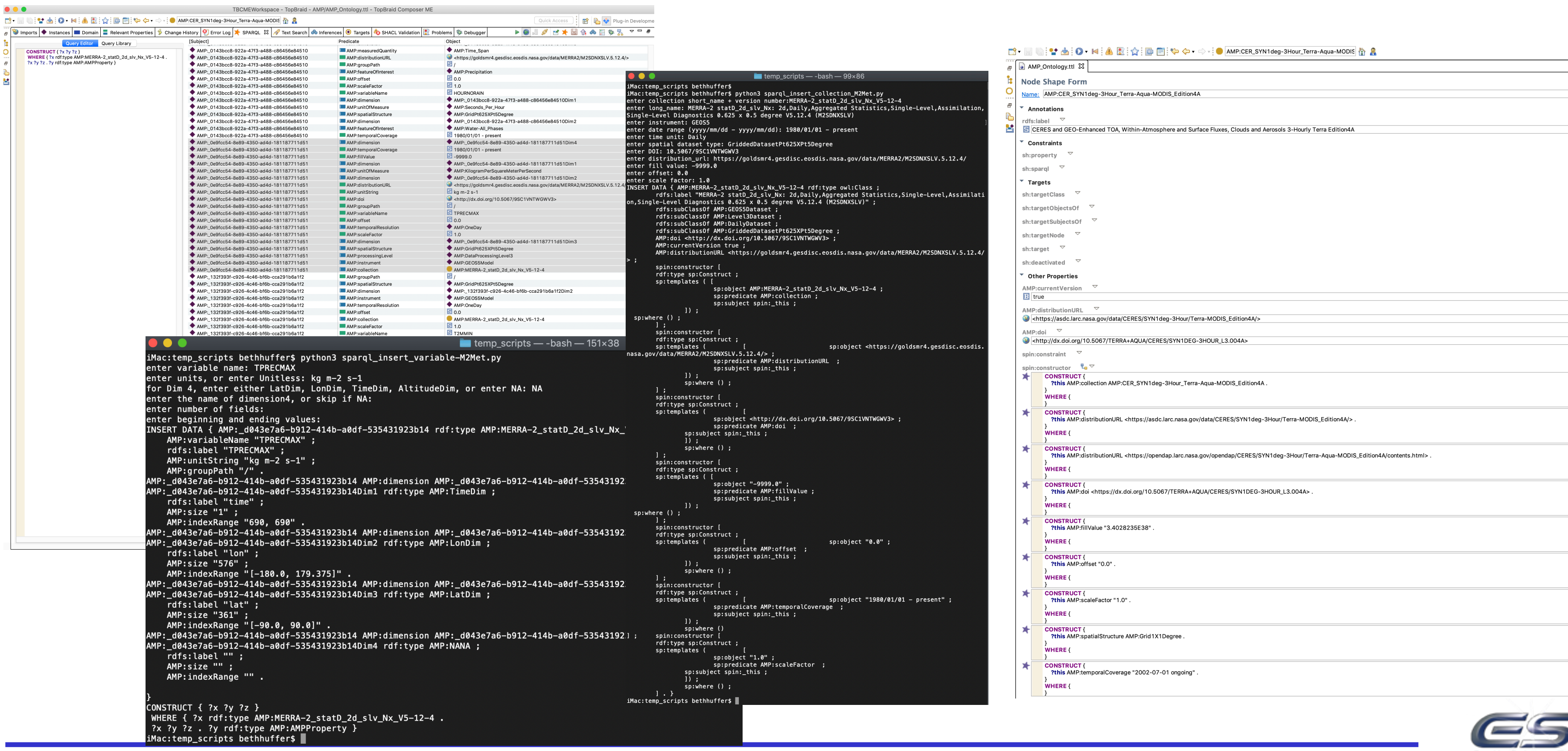
/lon	/lat	/time	/Albedo_inst	/AvgSurfT_inst
degrees_east	degrees_north	date	%	K
40.5	-14.5	2014-01-01	20.882565	300.65137
40.5	-13.5	2014-01-01	19.73688	299.9783
40.5	-12.5	2014-01-01	19.773048	299.9726
40.5	-2.5	2014-01-01	15.350428	303.07785
40.5	-1.5	2014-01-01	16.658976	303.86224
40.5	-0.5	2014-01-01	17.662725	304.30237
40.5	0.5	2014-01-01	17.940388	305.42383
40.5	1.5	2014-01-01	16.766033	305.63602
40.5	2.5	2014-01-01	15.799581	304.23944
40.5	3.5	2014-01-01	14.35454	302.46603

SD3P

Technical and Science Advancements

SEMANTIC ANNOTATOR

- A prototype semantic annotator uses semi-automated methods (scripts and inference rules) to generate metadata records;
- Created over 200 test records with “gold-standard” metadata records for sample radiation, air quality, and precipitation data products for training the CNN classifier



ESTO
Earth Science Technology Office

Technical and Science Advancements

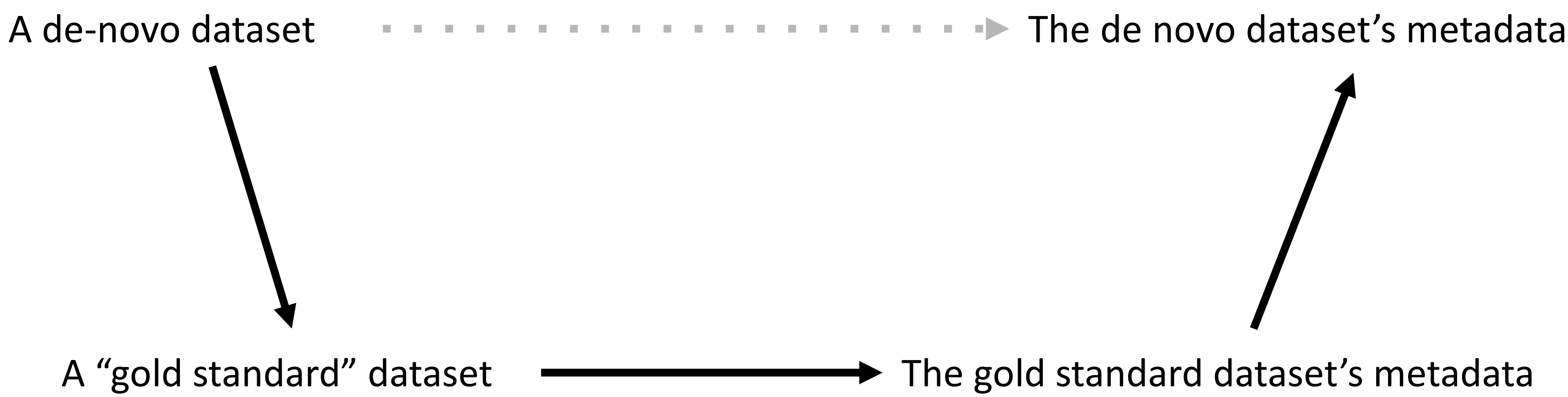
ML PIPELINE

The goal of the ML pipeline is to produce the portion of the metadata record that describes a dataset's science characteristics

A de-novo dataset $\longrightarrow$ The de novo dataset's metadata

ML PIPELINE

We determine the de novo dataset's science characteristics by comparing it to our gold standard datasets, and determining how similar the de novo dataset is to a gold standard



Given: a set of measurements through space and time

Can we: classify them by similarity to each of a set of well-known datasets (the *gold-standards*)

So that: we can infer metadata for a de-novo dataset

Technical and Science Advancements

THE OVERALL GOAL

We want to learn a mapping from de-novo datasets to discrete probability distributions over gold-standard datasets. These discrete distributions are embeddings that we can use to compute similarity.

de novo variable 1

g	Pr(g)
g_0	0.08
g_1	0.01
g_2	0.9
…	
g_n	0.01

de novo variable 2

g	Pr(g)
g_0	0.95
g_1	0
g_2	0.01
…	
g_n	0.01

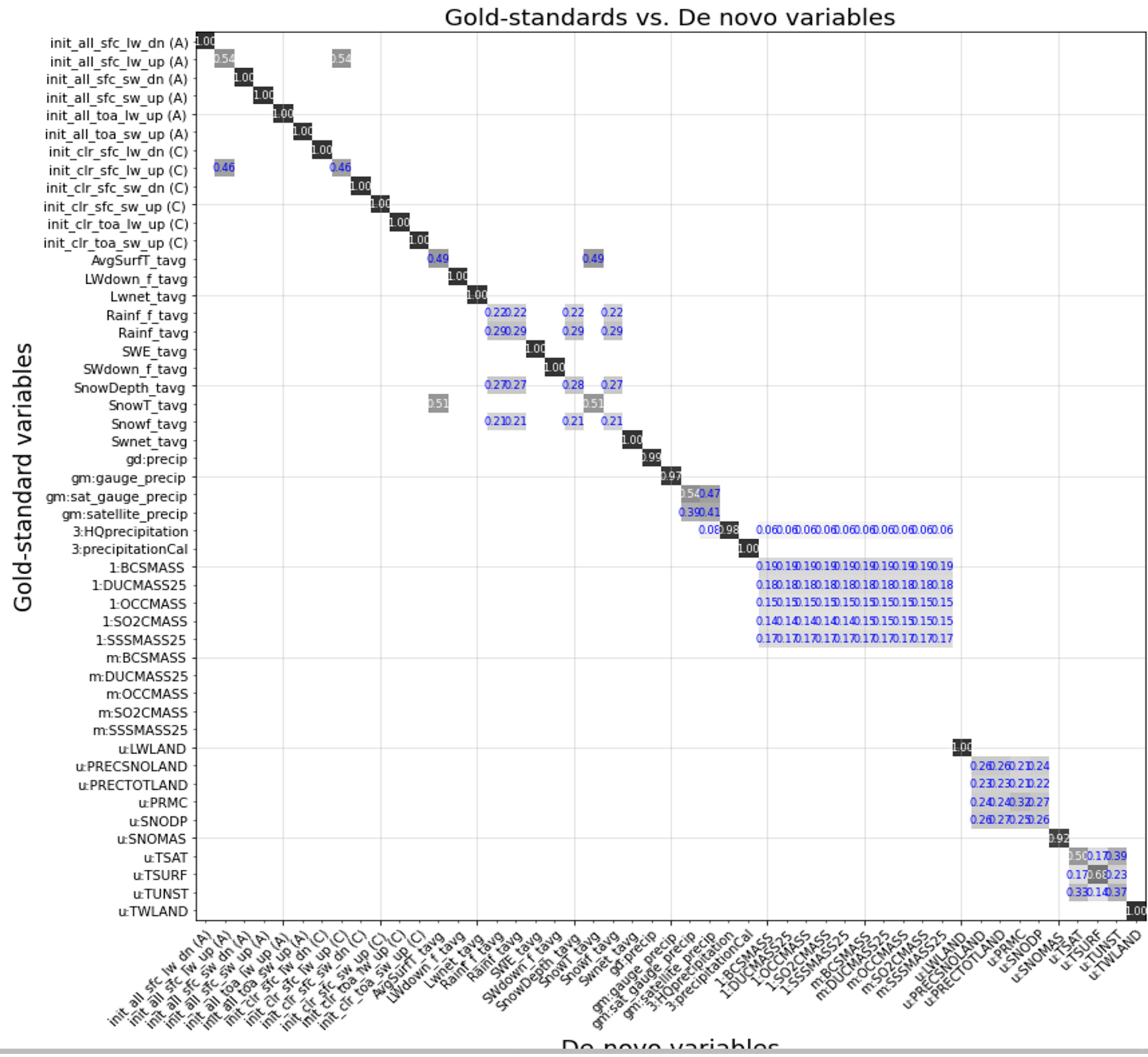
⋮

Technical and Science Advancements

Machine learning methodology

- Each dataset is a time-series of gridded measurements (L3)
  - Daily, monthly, and 1-hourly.
- Each temporal snapshot becomes a training example
- All examples are spatially resampled to $1.0° \times 1.0°$
- 8 collections, 49 variables -> 49 gold-standards
- 189,269 examples -> 50% training, 25% test, 25% hold-out
- Training:
  - Label is variable name
  - Input is 180×360 matrix
  - Output of model is 49 probabilities (one-hot, soft-max)
  - Standard convolutional network
- Inference:
  - Examples are same as in training
  - Combine distributions for examples into one distribution for a variable
- Interface to Semantic Annotator is probability distribution over gold-standards
  - Pick the most likely and use that, or
  - Use top-*k* most-likely gold-standards

Technical and Science Advancements



Presentation Contents

- Background and Objectives
- Technical and Science Advancements
- Summary of Accomplishments and Future Plans
- Publications - List of Acronyms

Summary of Accomplishments and Future Plans

- Prototypes of AMP components are complete, and AMP has advanced from TRL 3 to TRL 4
- Next steps include:
    - Building the end-to-end pipeline to connect the 3 components that currently operate independently
    - Writing APIs and testing AMP outputs with the ARIES system
    - Expanding the metadata repository with new metadata records
    - Analyzing and addressing weaknesses in the CNN

Presentation Contents

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

Thank you!

For questions, or for more information, please feel free to email us.

Beth Huffer beth@lingualogica.net
Simon Handley simon@lingualogica.net
Damian Gessler damian@lingualogica.net

List of Acronyms

• AI	Artificial Intelligence
• API	Application Programming Interface
• ML	Machine Learning
• SD3P	Science Data Pre-Processing Package