



Creating Satellite Data Products in the Cloud: SatCORPS Global Cloud Composites



Thad Chee¹, Louis Nguyen², William L. Smith, Jr.², David Painemal¹, Rabindra Palikonda³, Cecilia Fleeger³,
Konstantin Khlopenkov³, Andrei Vakhnin¹, Valeria Green⁴, and Audy J. Barnett⁴

¹ADNet, Inc., Hampton, VA, ²NASA-Langley Research Center, Hampton, VA, ³AMA, Hampton, VA, ⁴Booze Allen Hamilton, Inc, Norfolk, VA

Session: IN31B-0654

Creating Global Satellite Derived Cloud Products at NASA Langley

Real time satellite observations and real time derived cloud products are becoming an important tool for both science as well as business ventures. The SatCORPS group leverages public and private cloud-based sources of satellite observations to create its Global Cloud Composite product in near real-time. This dataset allows others access to cloud information that can be accessed directly from the cloud.

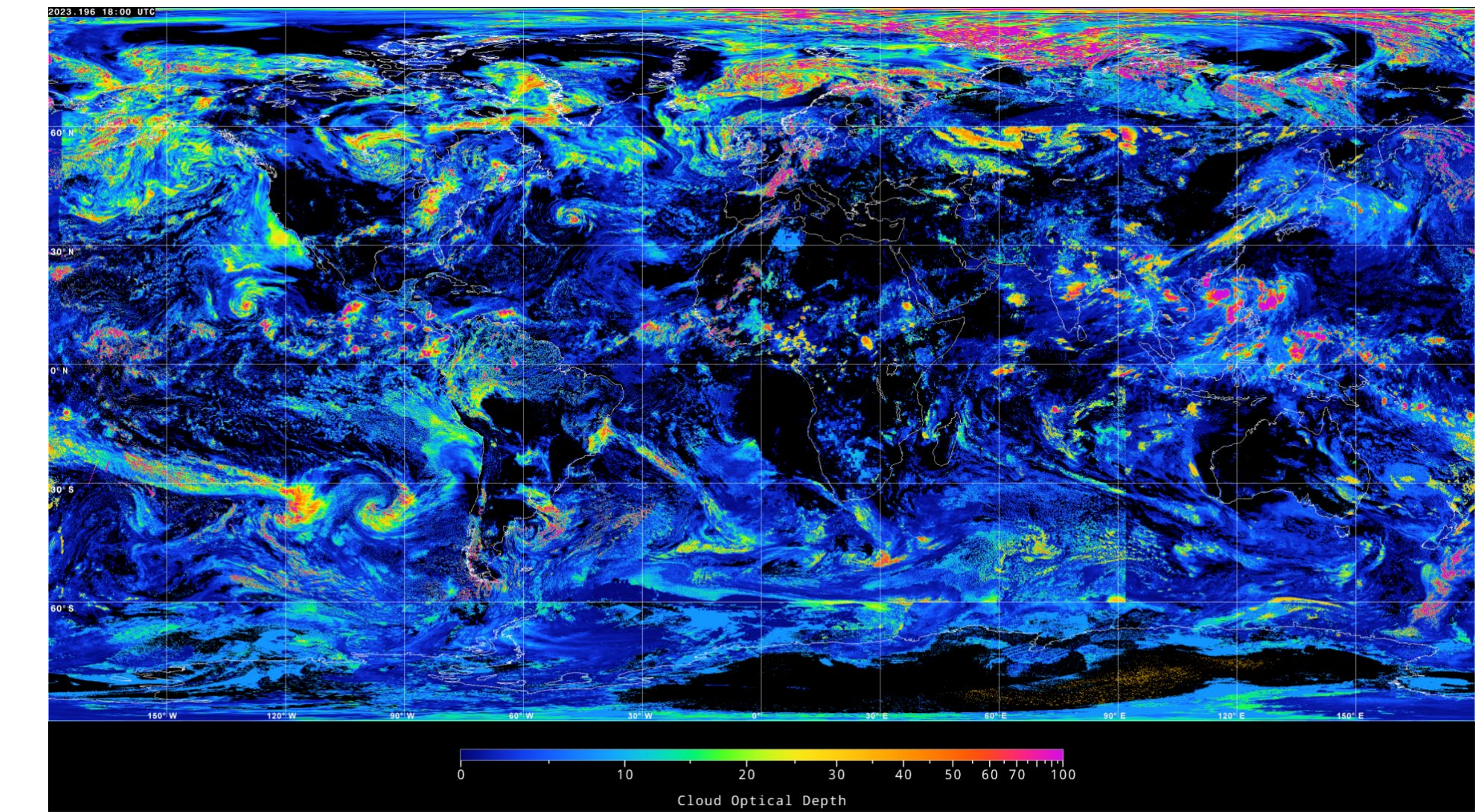
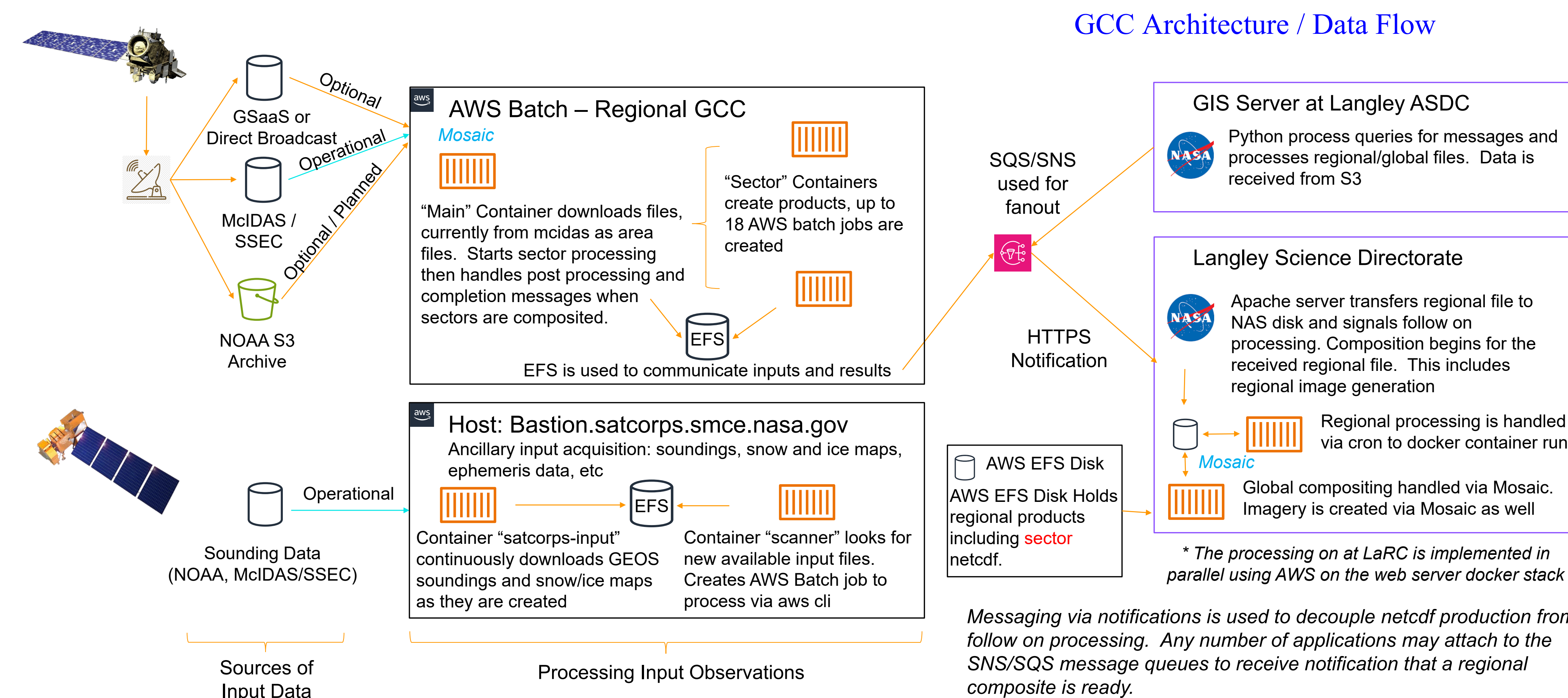
In this work, we describe the software algorithms and software infrastructure that we have created to create and distribute this product through our hybrid cloud and on-premises system that leverages the strengths and weaknesses of each platform.

We will also describe the GCC product itself in terms of the scientific parameters available, resolution and temporal availability. Finally, we will also describe our web-based tools including our Geographic Information System that can be used for visualization and analysis. The products from the processing can be geospatially-service enabled as ArcGIS Image Services and Open Geospatial Consortium (OGC) Web Mapping/Coverage Services for visualization and analysis via a web mapping application and common geospatial software.

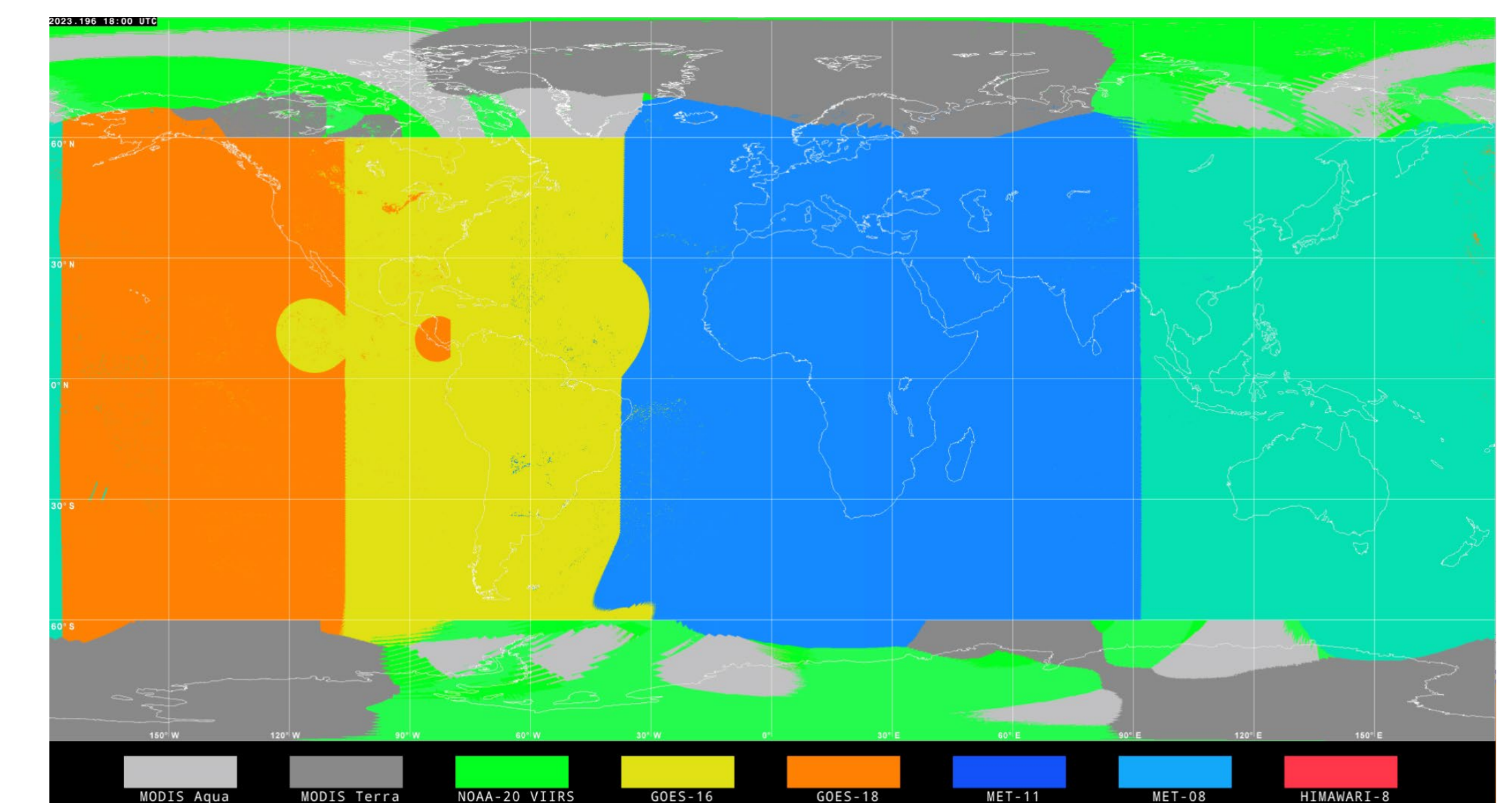
The SatCORPS Global Composite Cloud product provides sophisticated global cloud products with very low latency which we see that as filling a rapidly growing need in the research, modelling and business community. We present a detailed description of the SatCORPS GCC product in terms of the capabilities and research benefits as well as the delivery architecture in AWS. We take the discussion further and describe the web tools that the SatCORPS group has developed that allow users to access and use the GIS data that it creates.

Global Cloud Composite Cloud Products: Why and How?

The team considered a number of factors when designing a product meant for real-time distribution. One key element is that the products must be reliably delivered to be used in national weather models. After working with AWS, the system has proven itself to be a cost effective and reliable system. We leverage the various AWS services while also making sure similar functionality is available in other cloud platforms so that the system is not vendor locked. The services used are described in detail below and are: EC2, Batch, Lambda, SNS/SQS, S3 and EFS.



Supporting Architecture



Services and their contribution:

The services used within AWS accomplish different aspects of the product creation:

EC2 / AWS Batch: EC2 allows using hosts to serve the products, allow access to EFS and the AWS environment and to operate the batch processing system. AWS batch is used over clustering solutions because of the nature of the processing. Information processing is done in discrete batches (the nominal observation time) which leads to large slack periods that would otherwise need to be paid for with a cluster system. Batch allows massive scale up and tear down to provide compute resources when they are needed.

Lambda: Serverless computing allows for reliable and timely response to events.

SNS/SQS: SNS, with its fanout capability when used with SQS and HTTPS allows for distributed and parallelized processing. SQS allows for multiple service clients to service requests which allows distributed systems to be used at scale. Further, the decoupling of processing allows for CI/CD deployments where hosts can be removed for service/patching without bringing the system down.

MQ: MQ is used to provide real-time logging and is used with older systems.

EFS: Because EFS storage size is only limited by the ability to pay for the storage, this storage solution is well suited for production operation. A major source of downtime in prior systems the SatCORPS group has had is disk full conditions.

S3: While unlimited in storage size as well, S3 allows for holding information at much lower cost than EFS.

CloudFormation: This service is used to document and implement the design of the system. All resources are created from cloud formation templates which allows us to implement infrastructure as code.

GroundStation: AWS Groundstation is used when the an ultra-fast generation of data is needed. Information is captured from the satellites via direct broadcast (LEO) and can be injected into the production system allowing for disaster response information to be generated.

GitLab: Key to CI/CD is the git system which versions the entire production system as well as the infrastructure. Git allows the multiple developers on the team to work in a coordinated way.

Future Directions:

Going forward, we've identified a number of areas of research that would improve the system tremendously:

Archival capability: Over time, our science algorithms have improved in the quality of the products created. Since the system is operated in real time, products created by an earlier version of the programming is not made current. An archival capability for processing would allow the entire dataset to be uniform in terms of product quality, ensuring that improvements are in all products.

Machine Learning Algorithms: We are currently deploying a machine learning enhanced version of our software that improves results within the terminator region.

Multiple Region Deployment: AWS itself has gone offline a few times and by operating the system in multiple regions, we can mitigate this issue. Having the system as cloudformation templates is key to multi-region deployment.

Satellites Used to Create the Dataset

Satellite	Orbit	Region
GOES-18	Geostationary (GEO)	West Americas
GOES-16	Geostationary (GEO)	East Americas
METEOSAT	Geostationary (GEO)	Africa
METEOSAT	Geostationary (GEO)	India
MTSAT	Geostationary (GEO)	Eastern Pacific
NOAA-20	Polar Orbiting (LEO)	Globel
SNPP	Polar Orbiting (LEO)	Global
TERRA	Polar Orbiting (LEO)	Global
AQUA	Polar Orbiting (LEO)	Global

•Corresponding author email: Thad.L.Chee@nasa.gov

Cloud Products and Data Site: <https://satcorps.larc.nasa.gov/GCC-SNWG>

