Connecting Lake Observatories to Space-Based Missions: Global Lakes Ecological Observing Network (GLEON), NASA Surface Biology and Geology (SBG), and the Environmental Data Initiative (EDI)

Paul Hanson[1,1], Kathleen Weathers[2,2], Stephanie Schollaert Uz[3,3], Mark Servilla[4,4], and Corinna Gries[5,5]

[1]University of Wisconsin–Madison
[2]Cary Institute of Ecosystem Studies
[3]NASA Goddard Space Flight Center
[4]University of New Mexico Main Campus
[5]University of Wisconsin Madison

January 20, 2023

Abstract

Assessing and understanding the extent and trajectory of change in inland waters is a great challenge, due in part to both differing methods — and cultures — of agencies that provide synoptic observations of Earth's systems as well as the community of lake scientists whose research generates heterogeneous and distributed in situ data. Advancements require socio-technological initiatives that harness the resources of the highly diverse and distributed community of ecologists, as well as the products and expertise of the satellite remote sensing community. Here we describe a prototype for linking in situ and remotely sensed data for lakes through the collaborative efforts of the Global Lake Ecological Observatory Network (GLEON), the Environmental Data Initiative (EDI), and NASA. GLEON provides a community of lake scientists and data from lake observatories. EDI curates and publishes data and ensures conformity to rigorous FAIR principles. NASA provides the expertise and workflows to deliver remotely sensed data products on demand. The integration of the data and the communities provides a foundation for a new generation of lake science.

Connecting Lake Observatories to Space-Based Missions: Global Lakes Ecological Observing Network (GLEON), NASA Surface Biology and Geology (SBG), and the Environmental Data Initiative (EDI)

Paul Hanson, Kathleen Weathers, Stephanie Schollaert Uz, Austin Delany, Sean Bertalot, Cal Buelo, Carmen Cillero Castro, Corinna Gries, Mark Servilla

University of Wisconsin-Madison, Cary Institute of Ecosystem Studies, NASA GSFC, 3edata Environmental Engineering S.L., University of New Mexico

PRESENTED AT:

NASA

Surface Biology and Geology (SBG)

NASA has initiated a new study for the Surface Biology and Geology (SBG) Designated Observable, identified in the National Academies of Sciences, Engineering and Medicine (NASEM) 2017 Decadal Survey, “Thriving on Our Changing Planet: A Decadal Strategy for Earth Observations from Space.”

The Decadal Survey document presented a clear vision for the combined roles of visible to shortwave infrared imaging spectroscopy and multispectral or hyperspectral thermal infrared imagery in addressing terrestrial and aquatic ecosystems and other elements of biodiversity, geology, volcanoes, the water cycle and applied sciences topics relevant to many societal benefit areas.

The SBG Study is currently in the first phase of identifying as many feasible observing architectures as possible that achieve the Decadal Survey science objectives, including system concepts from the Hyperspectral Infrared Imager (HyspIRI) precursor study from the 2007 Decadal Survey, as well as new ideas and advances with instrument technologies. Candidate architectures will include SmallSat and medium class concepts, and industry and foreign partnerships.

A Research and Applications (R&A) team has been developed to include four working groups: algorithms, calibration and validation, applications, and modeling. A parallel team is envisioning candidate architectures for implementing observing system concepts, exploring capabilities from program of records and recent opportunities, and working from the input of the R&A team or their Science and Applications Traceability Matrix (SATM).

The objective of this study is to engage a broad science community and a variety of stakeholders to explore potential partnerships with commercial and international organizations.

SOURCES

GLEON: https://gleon.org/ (https://gleon.org/)

Workflow code:

NASA SBG Mission: https://science.nasa.gov/earth-science/decadal-sbg (https://science.nasa.gov/earth-science/decadal-sbg)

NASA Ocean Color: https://oceancolor.gsfc.nasa.gov/ (https://oceancolor.gsfc.nasa.gov/)

SeaDAS: https://seadas.gsfc.nasa.gov/ (https://seadas.gsfc.nasa.gov/)

EDI: https://environmentaldatainitiative.org/ (https://environmentaldatainitiative.org/)

EDI utilities for workflows: https://github.com/EDIorg/EDIutils (https://github.com/EDIorg/EDIutils)

Lake observational data package sources (script by Bertalot, Delany, 2021)

Buoy data doi: 10.6073/pasta/c03b39550e79d002d82a2281f8546c78

Manual sample doi:10.6073/pasta/f28e278afc34f1b7bd4f3cdc02b733a2 (edited)

Remote sensing data Sources (script by Buelo, 2021)

Level 1, 250 meter resolution data from MODIS Aqua and Terra were downloaded from the NASA Ocean Biology Processing Group (https://oceancolor.gsfc.nasa.gov/).

Data Processing: Data were processed using the SeaDAS remote sensing software from the NASA Ocean Biology Processing Group (https://seadas.gsfc.nasa.gov/). We used the main algorithm implemented in SeaDAS for calculating chlorophyll, the OCI algorithm, which stands for Ocean Color Index. It combines the OCx algorithm at high chlorophyll concentrations with the Color Index (Hu et al. 2012), which performs better at low chlorophyll concentrations. Both of these algorithms are based on the changes in relative reflectance at blue and green wavelengths as chlorophyll concentrations vary:

OCx algorithm: $\log_{10}(Chl) = a_o + \sum_{i=1}^{4} a_i * [\log_{10} \frac{R_{rs}(\lambda_{\text{blue}})}{R_{rs}(\lambda_{\text{green}})}]^i$, a_i = fit coefficients

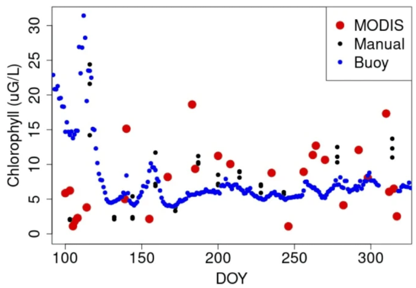
CI algorithm: $CI = R_{rs}(\lambda_{\text{green}}) - [R_{rs}(\lambda_{\text{blue}}) + \frac{\lambda_{\text{green}} - \lambda_{\text{blue}}}{\lambda_{\text{red}} - \lambda_{\text{blue}}} * (R_{rs}(\lambda_{\text{red}}) - R_{rs}(\lambda_{\text{blue}}))]$

$$Chl = 10^{-0.049 + 191.66 * CI}$$

Intermediate chlorophyll values between 0.15 and 0.2 mg/m3 use a weighted average of the two.

SeaDAS has a variety of methods for determining reflectance (and derived parameters) from sensor-measured radiance. We used the MUMM algorithm (Management Unit of the North Sea Mathematical Models) for atmospheric and aerosol correction. This method was developed to address the commonly used assumption that water-leaving NIR radiance is zero (Ruddick et al. 2000), which is often the case in clear ocean waters but not so in more turbid inland and coastal waters like those included in this study. Spatial averaging was enabled for cloudless scenes in the NIR and SWIR bands to compensate for the lower signal-to-noise ratio at longer wavelengths. Thresholds for clouds and chlorophyll concentrations were adjusted on a scene-by-scene basis.

LAKE WATER QUALITY: SENSORS TO SATELLITES



Lake chlorophyll concentrations from MODIS, *in situ* fluorometery (buoy), and manual samples for Lake Mendota, Wisconsin. Data are from 2014 and 2016 and standardized to manual samples.

Prototype automated workflow

A socio-technological solution connecting

GLEON with NASA's SBG mission

Remote sensing data are queried from NASA data repositories, combined with GLEON data, and visualized

Lake Mendota, Wisconsin
A GLEON lake observatory and study site of the North Temperate Lakes LTER

Water quality data from GLEON lake observatories are published in EDI

Water quality data are queried from EDI; high quality metadata and provenance tracking in a reliable repository ensure reproducible science

Difficult environmental problems, such as understanding global scale lake water quality dynamics, require interdisciplinary approaches. Collaborations catalyzed by NASA's SBG mission and that include GLEON, NASA, EDI, and the European Space Agency (ESA) are building the human and technical infrastructure required to sense and interpret water quality of inland waterbodies. These collaborations are working toward co-development of science that advances prediction and understanding of inland lake water quality.

This prototype demonstrates that workflows can easily be created to connect data from satellite remote sensing and lake observatories, through the EDI. The high quality metadata standards of EDI ensure lake observatory data products have persistent DOI identifiers, provenance tracking, and associated metadata that give credit to the data providers.

Lake Mendota, is a eutrophic lake in Madison, Wisconsin. As a centerpiece of the Madison metropolitan area, L. Mendota provides many services to visitors and the local community, including swimming, sailing, and fishing. Unfortunately, periodic harmful algal blooms sometimes render the lake unfit for recreation. Improved monitoring and understanding of lake water quality dynamics are high priorities for the Madison area community. The North Temperate Lakes Long Term Ecological Research (NTL LTER) program manages a the Lake Mendota buoy and runs a monitoring program that has collected decades of limnological data. NTL LTER provided the observational data for this demonstration.

GLOBAL LAKE ECOLOGICAL OBSERVATORY NETWORK (GLEON)

The Global Lake Ecological Observatory Network conducts innovative science by sharing and interpreting high-resolution sensor data to understand, predict and communicate the role and response of lakes in a changing global environment. GLEON creates a framework for collaboration through community policy, process, and structure. *See "Lake Expedition 2022" below for graduate student opportunity!*

A network of people

GLEON has >850 members, approximately ⅓ of whom are graduate students. Members represent 62 countries. GLEON, as a "learning organization", has evolved in its approach to data. Built-in mechanisms for listening to the needs of the community, empowering the community to make change, and experimentation & adaptation have driven GLEON evolution.

A network of lake observatories

GLEON members have observatory data from >150 lakes (120 buoys). There are 40 lake observatories that are part of GLEON, representing 6 continents and 55 countries.

A network of lake data

Data are first class objects that are assets of the members. GLEON data are accessed through collaborative science amongst members. Building data capacity means building science capacity, which requires community capacity. It starts with building a community that has trust and established principles and norms. It moves forward with engagement of members in science opportunities.

Graduate Fellows Training Program

GLEON is in it's fourth cohort of graduate student fellows. Fellows build a team (typically 10-12 students) who learn team science skills while taking on a big data problem to address a science question. Lake Expedition 2020 (right) is working with computer scientists to use "knowledge guided machine learning" to study long term lake area change from satellite remote sensing data.

Lake Expedition 2022

The GLEON Fellowship Program is recruiting 10-12 graduate student to work as a team on lake water quality and remote sensing. Exciting opportunity to collaborate with scientists from NASA, the European Space Agency, and other GLEON members. For more information, contact Kathleen Weathers or Paul Hanson.

ENVIRONMENTAL DATA INITIATIVE (EDI)

Environmental Data Initiative

- Create . Package . Archive . Discover . Reuse -

The Environmental Data Initiative (EDI) (NSF DBI #1931143 and #1931174): EDI enables new science by making FAIR environmental data publicly available. Building on 40+ years of data management experience and a decade of continuous operation, EDI has developed reliable, secure, and fast access to a large repository of environmental data that meet the highest standards of ecological metadata. Especially valuable are thousands of datasets of consistent long-term 'on-the-ground' observations documenting environmental change. Persistent Digital Object Identifiers (DOIs) enable citation of data for new uses, and detailed provenance information give credit where credit is due. Research programs are building automated workflows that take advantage of EDI's application program interface for data submission, as well as search and download of data for reuse. NSF-funding has made EDI the repository of choice for environmental biology, including the Long Term Ecological Research (LTER) Network and the National Ecological Observatory Network (NEON). EDI provides a range of end-user support and training to help scientists realize their goals, from consulting with large research programs and their information managers to 'hands-on' support for individuals and small scale environmental science. EDI supports any environmental scientist who wants their high quality FAIR data published and reused.

EDI and a new era of science: Connecting the past to the present through data enables exciting new science. Scientists may use EDI data to calibrate models and sensors, inform experimental design, give their observations a larger spatial and temporal context, and explore a range of scientific concepts across broad space and time domains. Melding the data life cycle with the scientific lifecycle promotes creative reuse of data while ensuring transparent and citable provenance of data products. EDI provides many examples for automatically adding, finding, and using data so that scientists who are data providers or data users can include EDI in scientific workflows, such as ecological forecasting or harmonization of different observation platforms.

EDI, LTER, and NEON: The infrastructure that underlies EDI was developed in close collaboration with scientists and information managers from the LTER program, and the EDI repository contains tens of thousands of LTER datasets as well as NASA data products. NEON chose EDI as their preferred data repository for research involving NEON resources, including products derived from using NEON data and raw data from custom deployments of relocatable and airborne sensor platforms.

EDI and security: EDI enforces both operational and physical security measures to protect the integrity and accessibility of all archived data. Compute and storage infrastructure is managed by the University of New Mexico's Center for Advanced Scientific Computing in a secure and environmentally controlled facility. Network connectivity to all critical services is restricted to a "need-to-access" basis, is actively monitored, and requires key-based authentication software for user verification. The repository implements strong versioning of data packages, and all data and metadata are immutable. Data integrity is maintained through random checksum verification of all data objects and multi-layer redundancy (LOCKSS), using both local near-line and off-line backups, along with daily long-term archives that are replicated in AWS Glacier.