

NSF Spatiotemporal Innovation Center

July 2020 Monthly Newsletter

Edited by Michelle Ly, Eden Brunner, Jacqueline Chen, Asad Yamin, Ruoshan Dong, Phil Yang, Ziyue Xu

Content provided by Phil Yang, Wendy Guan, Dexuan Sha, Tao Hu, Qian Liu, Elizabeth Goldbaum, Benjamin Lewis, Devika Kakkar, Danny Brooke

Designed by Genevieve Chin

August 11th, 2020

[Join Us Now!](#)

Inside This Issue

1. An Environmental Data Collection for COVID-19 Pandemic Research
2. Cloud Classification Project
3. Ocean Science to COVID-19
4. Spatiotemporal Analysis of Medical Resource Deficiencies in the U.S. Under COVID-19 Pandemic
5. Webinars on Modeling COVID-19 Pandemic: Resources, Methodology and Applications
6. Update: Evaluating OmniSci: Open Source GPU- powered SQL Database
7. Update: China Data Lab (CDL): Developing an Online Spatial Data Sharing and Management Platform
8. Update: 2C: Sustaining a Network for Dynamic Mapping of Secondary Cities
9. Update: Elevating Research Excellence with Data Repository and AI Ecosystem

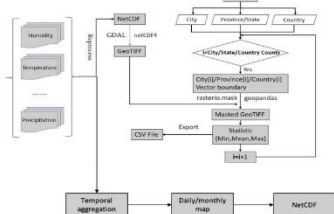


Figure 1. Workflow of Spatiotemporal Aggregation and Collocation.

An Environmental Data Collection for COVID-19 Pandemic Research

In response to the COVID-19 pandemic, the NSF Spatiotemporal Innovation Center created a spatiotemporal platform with various task forces to provide data support to a wide range of communities, including researchers, public citizens, and decision-makers in policy and health departments. Compared to other platforms that only utilize viral and health data, this platform, consisting of a team of international researchers, recognizes that virus-related environmental data collection (EDC) is also a key component for a broader understanding and geospatial analysis of the pandemic.

The EDC includes a range of environmental factors proven or with potential to influence and impacted by the spread and virulence of COVID-19 and its effects on human health such as humidity, temperature, air quality index and pollutants, precipitation, and nighttime light radiance (NTL). To ensure reliable and high-quality data publishing, researchers evaluated the collected environmental data in three dimensions: data integrity, consistency, and validity. The data was then processed and organized across multiple spatiotemporal scales for applications like a global mapping of daily temperature or a correlation of the COVID-19 pandemic to the mean values of climate and weather factors by city. By introducing raw input data, construction, and metadata of reprocessed data, data storage, and sharing and quality control methodologies, the NSF spatiotemporal platform seeks to serve as a data basis to help inform communities to mitigate and prevent the spread of COVID-19.

[Read More](#)

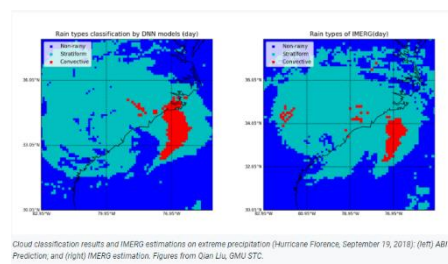
For further information and/or inquiries, email Ms. Qian Liu at qliu6@gmu.edu.

Cloud Classification Project

Collaboration between the NASA Center for Climate Simulation (NCCS) and George Mason University (GMU) has allowed for the fruition of many innovative machine learning projects under the GMU Center for Intelligent Spatial Computing, which is led by Director Phil Yang. One such project is the Cloud Classification Project, which provides more accurate, real-time detection of rainy clouds through a new deep learning method. The goal of the Cloud Classification Project is to classify clouds based on their potential to produce precipitation and their forms, positions, and other features. In collaboration with the NASA Langley Distributed Active Archive Center, researchers are investigating how to classify clouds into eight categories through a deep neural network method using various parameters such as brightness temperature differences, reflectances, optical depth, etc. With more data, the automatic cloud classification system can be used for real-time forecasting, decision-making support in rainfall-related disasters, and other weather-dependent events. The deep neural network was found to be more accurate than traditional machine learning methods. This project has discovered the advantages of using AI for modeling atmospheric phenomena, and more work with AI remains to be done in this field.

[Read More](#)

For further information and/or inquiries, email Dr. Chaowei Phil Yang at cyang3@gmu.edu.



Cloud classification results and IMERG estimates on extreme precipitation (Hurricane Florence, September 19, 2018): (left) AI Prediction, and (right) IMERG estimation. Figures from Qian Liu, GMU STC.



Ocean Science to COVID-19

In 2015, Chaowei Phil Yang, a professor at George Mason University in Fairfax, Virginia, started developing a novel Google-like search portal to allow scientists to effectively discover oceanographic data collected by satellites, airplanes and water-based sensors. However, what began as an effort to help scientists study the ocean slowly evolved into an effort to tackle the current COVID-19 pandemic. The portal, called “Mining and Utilizing Dataset Relevancy from Oceanographic Dataset” (MUDROD), was funded by NASA’s Earth Science Technology Office and made available to the public in 2017. MUDROD uses machine learning algorithms to understand and respond to queries about the ocean. When COVID-19 cases began to rapidly spread around the world, Dr. Yang used MUDROD’s algorithms to analyze the growing amount of data about the deadly virus. Together, he and his team updated MUDROD’s algorithm for the COVID-19 Spatiotemporal Rapid Response Gateway to locate information and trends from COVID-19 data and news stories by helping users understand real-time death and infection rates, responses to quarantine restrictions in different countries, adjustments in travel patterns in response to government guidelines, company hiring and layoff rates, and air quality changes during the winter, spring and summer of 2020. Thanks to the easy access to information MUDROD has provided, the web portal has helped lead to numerous publications in journals and at conferences

[Read More](#)

For further information and/or inquiries, email Dr. Elizabeth Goldbaum at elizabeth.f.goldbaum@nasa.gov or Dr. Chaowei Phil Yang at cyang3@gmu.edu.



Fig 3. Spatiotemporal visualization interface based on ArcGIS Dashboards.

Spatiotemporal Analysis of Medical Resource Deficiencies in the U.S. Under COVID-19 Pandemic

To estimate the shortness of medical resources, Mr. Sha of the STC GMU site developed an algorithm and relevant index to estimate medical burden at a county level from February 15, 2020, to May 1, 2020. To do this, hospital beds and critical care staff (CCS) were quantified to develop the Medical Resource Deficiency Index (MRDI) and Local Daily Medical Resource Deficiency Index (MRDId). The MRDI index represents the number of accumulated active confirmed cases normalized by the local maximum potential medical resources (total licensed beds and total CCS). Meanwhile, the MRDId represents the local daily medical burden, or the number of hospitalized patients that can be supported per ICU beds per CCS. An MRDId greater than one means local medical resources cannot fully support the hospitalized critically ill patients, or that the local medical burden is heavy. An MRDId less than one means that local medical resources are sufficient. The researchers used ArcGIS to visualize medical resource burden across the United States. According to their visualizations, general medical burden shifted from the east coast of the U.S. to midwestern states, and it was observed that Louisiana, Mississippi, Georgia, Tennessee, Indiana, and Nebraska are seeing a new wave of medical resource deficiencies. The MRDI and MRDId indices can be implemented into epidemic models down the road to estimate the severity of the pandemic or predict the outbreak cases in the U.S. and other countries.

[Read More](#)

For further information and/or inquiries, email Mr. Dexuan Sha at dsha@gmu.edu.

COVID-19 Data Challenges and Objectives**Challenges**

- Fragmented information
- Incomparable and lack of standard
- Separate from maps
- Temporary data connection
- Repeated work with low efficiency

Objectives

- Integration of data from different sources
- Standardization of data formats
- Integration of data and maps
- Permanent collections
- Improvement with efficient technology
- Collaboration on COVID-19 studies

Webinars on Modeling COVID-19 Pandemic: Resources, Methodology and Applications

China Data Lab (CDL) is a cloud-based geospatial data analysis platform for geospatial data gathering, management, analysis, visualization, and sharing. The experts participating within this forum include 7 review experts and 4 moderators, and the number of webinar registrations totaled to 1,333 individuals and approximately 1,985 attendees. The objectives of this platform include data sharing, workflow/research sharing, tool sharing and dissemination in an effort to unify and evaluate the information communicated between professionals across the world. The CDL created a series of webinars that compiled reviews on the topic of COVID-19 models and data collection. These webinars discussed methodology, applications, and internationally comparative studies on COVID-19 modeling. The webinar includes power points, recorded presentations, and lectures. One of the presentations, “An Overview of Research Data for COVID-19 Modeling,” was given by a group of specialists collaborating from the China Data Institute, Center for Geographic Analysis at Harvard University, Geocomputation Center for Social Science at Wuhan University, and the Spatiotemporal Innovation Center. In this presentation, the team delved into various topics surrounding the collection of research data for COVID-19 modeling such as data collection and integration, accessibility of resources for COVID-19, methodology, technology, research, and publication. All webinar videos and report power points can be accessed and downloaded from the Harvard Dataverse website.

[Read More](#)

For further information and/or inquiries, email Dr. Tao Hu at taohu@fas.harvard.edu

Update: Evaluating OmniSci: Open Source GPU- powered SQL Database

OmniSci is a high-performance database developed to harness the massive parallelism of CPU and GPU hardware. It can query up to billions of rows in milliseconds and is capable of high ingestion speeds. CGA has successfully completed the installation of OmniSci as a public app on the Harvard High Performance Computing cluster. Any Harvard researcher can now benefit from it by running interactive big data analysis cost effectively on large instances. The system user manual and installation scripts have been made available on CGA's GitHub for reference to all Harvard users, as well as users from other universities who would like to deploy the OmniSci app on their own computational clusters. Through several demos, we showed how the system can be used to answer questions related to COVID-19 research in real-time. In addition, we successfully completed the first use case of the Voter Affiliation Project that investigated the effect of geography on partisanship in the United States. The full computation for this project was performed on Harvard's Cluster instead of AWS, thus resulting in substantial cost savings for Harvard researchers. Additionally, we have made our Geo-tweets archive available to researchers through the Harvard cluster and provided open-source scripts to analyze this data using OmniSci. [Read More](#)

For further information and/or inquiries, email Ms. Devika Kakkar at devikakakkar29@gmail.com.

Update: Spatial Data Lab (SDL): Developing an Online Spatial Data Sharing and Management Platform

Spatial Data Lab (SDL) is an online platform that has the capabilities to generate, aggregate, and share spatiotemporal data, analytical tools, and cases. The platform is providing standardized datasets, completed workflow, enhanced methodology, evaluated results, and training resources for collaborating researchers to grasp the mechanism of the platform as quickly and efficiently as possible. In July 2020, the CDL team relayed effective methodology, applications, and international comparative studies on COVID-19 modeling through the webinar "Modeling COVID-19 Pandemic: Resources, Methodology and Applications." Within the series of webinars, Dr. Guan and Dr. Hu provided a presentation on "An Overview of Research Data for COVID-19 Modeling" on July 30, 2020. Furthermore, the CDL team began to apply the workflow tools on the SDL platform to replicate the compiled COVID-19 prediction models. [Read More](#)

For further information and/or inquiries, email Dr. Wendy Guan at wguan@cga.harvard.edu or Dr. Tao Hu at taohu@fas.harvard.edu

Update: 2C: Sustaining a Network for Dynamic Mapping of Secondary Cities

Secondary cities are non-primary cities, characterized by population size, function, and/or economic status. They are urban centers of governance, logistics, and production, and they often lack much data. The purpose of the 2C project is to address the geospatial needs of secondary cities. This project is facilitated in partnership with local organizations, and it focuses on applied geography such as human geography and thematic areas. By recording this data, the 2C project will enhance emergency preparedness, human security, and resilience, which are useful for helping fight the COVID-19 pandemic. As of now, the project is ready to migrate the 2C database from GeoNode to ArcGIS Online.

For further information and/or inquiries, email Mr. Benjamin Lewis at blewis@cga.harvard.edu or Dr. Wendy Guan at wguan@cga.harvard.edu.

Update: Elevating Research Excellence with Data Repository and AI Ecosystem

The Harvard Dataverse uses API endpoints that help provide statistics on dataset views, downloads, and use, which are utilized by the RMDS team in data recommendations. The Dataverse is an open source data repository platform where users can share, preserve, cite, and explore research data. RMDS Lab is a startup company helping develop transformative technologies for research with big data and AI. The project aims to establish a collaboration between the two teams and platforms to support worldwide scholars in data-driven research. By applying AI technology to evaluate data science studies, provide measurable references for data scientists on the accuracy, impactfulness, replicability, applicability, and other merit scores of data science case studies, and promote high-quality data science research through platform development, data sharing, community building and user training, the project can allow for the exploration of different solutions to help combat the current pandemic. The collaborative efforts between RMDS Lab, Harvard Dataverse and the 60 Dataverse installations worldwide, have strengthened immensely due to the COVID-19 pandemic, and it is expected to increase in the future. [Read More](#)

For further information and/or inquiries, email Mr. Danny Brooke at dannybrooke@g.harvard.edu.