

Utah State University

DigitalCommons@USU

Funded Research Records

Data Services

6-3-2022

Collaborative Research: Surface-Specific Aerosol Chemistry: Direct Observations, Kinetics, and Environmental Impact

Yi Rao

Utah State University, yi.rao@usu.edu

Follow this and additional works at: https://digitalcommons.usu.edu/funded_research_data



Part of the [Biochemistry Commons](#), and the [Chemistry Commons](#)

Recommended Citation

Rao, Yi, "Collaborative Research: Surface-Specific Aerosol Chemistry: Direct Observations, Kinetics, and Environmental Impact" (2022). *Funded Research Records*. Paper 228.

https://digitalcommons.usu.edu/funded_research_data/228

This Grant Record is brought to you for free and open access by the Data Services at DigitalCommons@USU. It has been accepted for inclusion in Funded Research Records by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



12 Data Management Plan

(1) Data Acquisition: The acquisition of NASA mission data and metadata, as well as those from Ground-Based Observatories (GBOs), has been highly simplified due to the development of multiple public repositories in the last couple of years. For this project, we plan to use data the following repositories containing NASA mission data and metadata. (i) GOES X-ray flux data (<https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/>). (ii) GOES Proton and Electron Flux data (<https://satdat.ngdc.noaa.gov/sem/goes/data/full/>), (iii) Virtual Solar Observatory (VSO): <https://virtuelsolar.org/>, for SOHO/COSTEP particle fluxes, and (iv) ACE/EPAM particle fluxes (<http://www.srl.caltech.edu/ACE/ASC/DATA/level3/epam/>). For SEP events catalogue, we will utilize the catalogue from Papaioannou et al. [55] and the online archives of Coordinated Data Analysis Workshops (CDAW) (https://cdaw.gsfc.nasa.gov/CME_list/sepe).

(2) Dissemination of Data Products: All the data used and generated during this project will be made publicly available in a timely manner on the Web (via our own project websites, as well as the publicly-available data portals), properly formatted and documented. We want to publish our data products on global data portals: (i) we expect computer scientists, statisticians, data scientists and space weather researchers to be interested in our data products as dataset benchmarks for spatio-temporal data mining. Therefore we will upload our data to the UC Irvine Machine Learning Repository (<http://archive.ics.uci.edu/ml/>), the StatLib Datasets Archive at Carnegie Mellon University (<ftp://rcom.univie.ac.at/mirrors/lib.stat.cmu.edu/datasets/.index.html>), and the Dataverse repository at CfA (<https://dataverse.harvard.edu/>); (ii) we expect solar physics and space weather scientists to use our data for knowledge verification purposes as well as for the development of new scientific hypotheses. Therefore we will disseminate them even further via the VSO and public data sharing services, such as Zenodo (<https://zenodo.org/>). We have used all of these repositories multiple times in the past and are familiar with the data uploading, downloading and sharing practices.

(3) Plans for Archiving, Preservation and Access to Data: All our data products will be retained for at least five years after acceptance of publications. We intend to keep this data for much longer, making it available for further benchmark evaluations and research and comparative analyses with other methods and with other/future data sets. The research team will also work with the respective library services at Utah State University for long-term data curation and preservation. The server will leverage common fault tolerance techniques (e.g., RAID, periodic backups, etc.) to ensure that critical data loss can be mitigated. The department runs a state-of-the-art automated backup infrastructure that has already been expanded to meet the needs of this project. USU is committed to preserve the data and resources beyond the end of the funding period, ensuring continued accessibility to all the resources developed by this project. USU will support the costs of maintaining the network, the servers, and the backup infrastructure.

(4) File Formats for Broad Data Sharing: All the released data products will be available in three formats: (i) text with meta-data information that characterizes all fields in our data records - this is intended to allow easy data transformations between different operating systems and programming toolkits, (ii) the spatio-temporal database formats using OpenGIS (www.opengeospatial.org/) standards - this is intended to facilitate computer science research and easy data fetching to spatial databases, and (3) in data formats popular in specific research domains, for example, HEK-compatible XML files, popular among solar physicists.

(5) Policies for Reuse, Redistribution, and Derivative Production: All the products developed by the activities of this project will be disseminated to the public through open repositories and made available for reuse, reproduction, and modification with the least amount of restrictions. All the software, tools, and algorithms will be distributed using licensing practices (e.g., GNU General Public License) that reflect the principles of open source, which include source code availability, allowing redistribution and modification (to be redistributed under same terms), non-discrimination against persons, groups, or fields of endeavor, and non-contamination of other software products. Educational and outreach materials will be freely distributed and made available under similar licensing schemas, encouraging reuse, modifications, and redistribution. If any discoveries or inventions are directly connected with these products, we will grant access to such products once appropriate invention disclosures and/or provisional patent filings are made, while insuring that individual identity is not compromised under any circumstances.

(6) Policies for Access and Sharing: All the data resulting from this project will be made available to teachers, students, practitioners, and the broader audience. These data will be administered in accordance with USU policies and NSF Data Management Guidelines pertaining to intellectual property, record retention, and data management. Different data will be made available at different stages of the project. They will either be published through the project website or disseminated upon request, depending on the popularity and sensitivity of those data, as well as related privacy and security issues. In more details,

- (a) Research documents: Published research articles will have copyright protection. However, these articles will also be shared through the project website when their publishers allow disseminating papers for personal research use. Technical reports will be shared through the project website.
- (b) Research software and datasets: Implementation of research algorithms, synthetic data sets, and pre-processed real data sets will be packaged for distributions at the project website, representing complete pieces of knowledge. The project website will be designed to record each entry that downloads such packages.
- (c) Education and outreach products: Such materials including education modules will be made available to the public through the project website once they are ready for use.

The PIs will work closely with the legal offices at USU to develop the most effective policies for access and sharing. All the packaged materials will be published through the project website as long as the privacy, confidentiality, security, intellectual property, and copyrights are still protected. The project website will be implemented to enforce access policies, control, and data integrity and security.

(7) Note on Data Privacy Concerns: Since all aspects of our input data sets are publicly available on the web, there are no privacy or confidentiality implications both for the Data Inputs, which we plan to use during our project, as well as Data Products generated by our research.