Big Data and Data Fusion: Exploring EPA's Air Quality Products with Google Earth Engine

Elizabeth Mannshardt and Barron Henderson
EPA

The EPA's Air Quality Assessment Division's (AQAD) Data Fusion Team (DFT) provides leadership and direction on the characterization of air quality through development of emissions, monitoring, modeling, and data analysis. AQAD has collaborated with the research community in the development and application of statistical techniques that combine ambient monitoring data with air quality model results to characterize pollutant concentrations for use in various policy and regulatory assessments. These techniques are typically referred to as "data fusion", with "downscaler" a specific method currently utilized at EPA. Data fusion can overcome the spatial limitations of monitoring networks and benefit from the spatial and temporal coverage of air quality modeling (CMAQ). There are various types of techniques that have been created for and are applied to EPA projects. These range from informing the estimation of public health and environmental welfare benefits of regulatory programs to characterizing pollutant concentrations for use in exposure modeling or related health and environmental welfare studies.

A main component of the mission of EPA's Data Fusion Team is to identify and understand current data fusion methods as well as new or emerging application areas of data fusion methods and tools. In a world of ever-increasing big data, the need is even more prevalent for efficient processing and mathematical modeling techniques - as well as data visualization tools – for massive amounts of air quality data on vast spatial and temporal scales. Cloud-based platforms such as Google Earth Engine (GEE) "combine a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities". The public geospatial data archive includes more than thirty years of historical imagery and scientific datasets, updated and expanded daily, as well as a web-based code editor for interactive and fast algorithm development. The Data Fusion Team is interested in a pilot study exploring GEE's platform and functionalities to evaluate at remotely-sensed air quality products and indicators, and compare them with EPA's downscaled data fusion air quality products, modeled output, and monitored pollutants. An application in air quality will also be developed to showcase the power of big data, data fusion, and data visualization.

Abstract Information (continued)

The ultimate goal of this project is to explore GEE as a framework for fusing and evaluating satellite products and ground-truth monitoring station data, with intention of peer-review publications that communicate the utilization, evaluations, and applications of data fusion methods with modeled, monitored, and satellite products on a cloud-based big-data platform.

References:

[1] Berrocal, V. J., Gelfand, A. E. and Holland, D. M. (2010a). A spatio-temporal downscaler for outputs from numerical models. J. Agric. Biol. Environ. Stat. 15 176–197.doi:10.1007/s13253-009-0004-z

[2] https://www.epa.gov/air-research/community-multi-scale-air-quality-cmaq-modeling-system-air-quality-management

[3] https://earthengine.google.com/

[4] https://earthengine.google.com/