Public Health Data Applications Using the CDC Tracking Network: Augmenting Environmental Hazard Information with Lower-latency NASA Data

H.M. Amos[1,2,*], N.K. Skaff[3], S. Schollaert Uz[1], F.S. Policelli[1], D. Slayback[1,2], E. Macorps[1,4], M.J. Jo[1,5], K. Patel[2,6], C.A. Keller[1,7], P. Abue[2,6], V. Buchard[1,5], A.K. Werner[3]

[1]Earth Science Division, Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD, USA; [2]Science Systems and Applications, Inc., Lanham, MD, USA; [3]National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, USA; [4]NASA Postdoctoral Program; [5]University of Maryland Baltimore County, Baltimore, MD, USA; [6]University of Texas, Austin, Austin, TX, USA; [7]Morgan State University, Baltimore, MD, USA

*Corresponding author: Helen M. Amos, helen.m.amos@nasa.gov

ORCiD: Amos 0000-0002-0297-6643, Skaff: 0000-0002-5929-3966, Schollaert Uz: 0000-0002-0937-1487, Policelli: 0000-0002-7446-7338, Slayback: 0000-0002-4291-843X, Macorps: 0000-0003-4307-8978, Jo: 0000-0002-4351-2167, Patel: 0000-0002-2368-8624, Keller: 0000-0002-0552-4298, Abue: 0000-0001-6689-2978, Buchard: 0000-0001-9433-9205, Werner: 0000-0001-5838-1254

Abstract

Exposure to environmental hazards is an important determinant of health, and the frequency and severity of exposures is expected to be impacted by climate change. Through a partnership with the U.S. National Aeronautics and Space Administration, the U.S. Centers for Disease Control and Prevention's National Environmental Public Health Tracking Network is integrating timely observations and model data of priority environmental hazards into its publicly accessible Data Explorer (https://ephtracking.cdc.gov/DataExplorer/). Newly integrated datasets over the contiguous U.S. (CONUS) include: daily 5-day forecasts of air quality based on the Goddard Earth Observing System Composition Forecast (GEOS-CF), daily historical (1980-present) concentrations of speciated PM2.5 based on the Modern Era Retrospective analysis for Research and Applications, version 2 (MERRA-2), and Moderate Resolution Imaging Spectroradiometer (MODIS) daily near real-time maps of flooding (MCDWD). Data integrated into the CDC Tracking Network are broadly intended to improve community health through action by informing both research and early warning activities, including (1) describing temporal and spatial trends in disease and potential environmental exposures, (2) identifying populations most affected, (3) generating hypotheses about associations between health and environmental exposures, and (4) developing, guiding, and assessing environmental public health policies and interventions aimed at reducing or eliminating health outcomes associated with environmental factors.

Plain Language Summary

The well-being of people and communities can be impacted by hazards like polluted air and flooding. This paper describes a new partnership to bring more timely information about U.S. air quality and flooding into a centralized online environmental public health platform, the CDC Tracking Network's Data Explorer. Timely monitoring and forecasting of environmental hazards have the potential to reduce harmful exposures by enabling early warnings and other timely public health interventions. Currently, near real-time data may not be readily accessible to public health practitioners, particularly those that need to react to public health emergencies, because the data are fragmented across multiple systems in formats inconsistent with typical public health uses. Developing data pipelines and processes to host this data in a usable and standardized format and disseminating the data via an existing online platform assists in making important information about environmental health hazards more readily accessible to public health practitioners, decision makers, and partners.

Key Points

- Timely information can be critical for public health decision making, but environmental hazard data can be fragmented or hard to use.
- This work brings timely, routinely updated NASA data into CDC Tracking Network's Data Explorer in familiar and accessible formats.
- For CONUS, daily NASA air quality forecasts, historical air quality data, and daily flood maps are integrated into the Data Explorer.

1. Introduction

Harmful environmental exposures are a significant contributor to morbidity and mortality in the U.S. and numerous health outcomes are expected to worsen under climate change [USGCRP, 2016; Watts et al., 2021]. The U.S. National Aeronautics and Space Administration (NASA) uses satellites, aircraft, ground measurements, and Earth system models to study the long-term state of the environment as well as produce near real-time (NRT) data about environmental hazards such as natural disasters [NASA, 2023a]. The U.S. Centers for Disease Control and Prevention's (CDC) National Environmental Public Health Tracking Program (CDC Tracking Program) tracks exposures and health effects associated with environmental hazards [CDC, 2023a]. The Environmental Public Health Tracking Network (CDC Tracking Network), built by the CDC Tracking Program, is a dynamic, web-based system that brings together health data and environmental data from a network of partners to provide timely, relevant information for better community health (https://ephtracking.cdc.gov/). This paper describes recent efforts to augment information provided through the CDC Tracking Network with lower-latency, higher spatiotemporal resolution NASA datasets for high priority environmental hazards. The overall goal is to enable health researchers and practitioners to access critical environmental data needed to understand and respond to health risks and make data-informed decisions.

The mission of the CDC Tracking Program is to provide information from a nationwide network of integrated health and environmental data that drives actions to improve community health [CDC, 2023a]. The CDC Tracking Network is a multi-tiered, web-based system with components at national, state, and local levels that unifies health and environmental data from a network of varied sources and makes that information publicly available to its audience of public health researchers, professionals, decision makers, and the public in standardized formats [CDC, 2023a]. In collaboration with federal, state, and local partners, priority environmental health issues and key surveillance questions are identified. Existing data are evaluated for their ability to inform these issues and then integrated into the CDC Tracking Network. Data are disseminated through the CDC Tracking Network's flagship product, the Data Explorer (https://ephtracking.cdc.gov/DataExplorer/), as well as dashboards, infographics, and an application programming interface (API). Additionally, when gaps or new data needs are identified, the CDC Tracking Program collaborates with partners (https://www.cdc.gov/nceh/tracking/partners) to develop standards for data collection, develop models, expand the utility of non-traditional public health data, or develop new methodologies for using existing data.

Near real-time monitoring and forecasting of environmental hazards have the potential to reduce harmful exposures by enabling early warning systems and other timely public health interventions [e.g., WMO, 2023]. Currently, critical near real-time data may not be readily accessible to public health practitioners, particularly those that need to react to public health emergencies, because the data are fragmented across multiple data systems and are not formatted in a manner consistent with typical public health uses (e.g., using administrative boundaries such as county and census tract) [Liu et al, 2021]. Augmenting and improving data pipelines and processes to host these data in a usable and standardized format and disseminating these data via the CDC Tracking Network will assist in making important information about environmental hazards more readily accessible to public health practitioners and partners, including relevant CDC programs, and state, tribal, local, and territorial health departments.

The objective of this work is to leverage the NASA Goddard Space Flight Center (GSFC) expertise in processing and interpreting a wide range of environmental satellite data products and to make these data more accessible to public health practitioners through the CDC Tracking Network. Section 2 describes the data needs of the CDC Tracking Program and efforts to identify environmental datasets by NASA GSFC to meet these needs. Section 3 details the data systems methodology and implementation processes to define and establish routine workflows for transforming and transferring data in an automated and timely fashion. Section 4 illustrates applications of the data. Section 5 summarizes limitations of the datasets and offers some guidance on appropriate uses in public health context. Sections 6 and 7 capture outlooks and conclusions, respectively.

2. Public Health Environmental Data Needs

This section describes the collaborative process between the CDC Tracking Program and NASA GSFC around the use of available environmental hazard data to address current CDC data needs. Priority topic areas include air quality, flooding, extreme weather, wildfires, and climate impacts. These priorities reflect gaps in existing content on the CDC Tracking Network's Data Explorer, feedback from the CDC Tracking Network's Content Work Group focused on air quality, climate, and weather as areas of programmatic emphasis within the CDC National Center for Environmental Health and the CDC Division of Environmental Health Science and Practice. The CDC Tracking Program identified a need for lower latency data (i.e., shorter lag time between data production and availability to users) because of the potential to inform public health emergencies and to have timelier data for decision-making. This paper presents progress on the first two priority topic areas of air quality and flooding.

2.1 Data selection

NASA GSFC produces numerous publicly available data products relevant for air quality (https://airquality.gsfc.nasa.gov/) and flooding [NASA, 2023b]. Product applicability was assessed based on geographic coverage and resolution, temporal coverage and resolution, latency, means of data access, file formats, and levels of processing (e.g., Level 1–Level 5 [NASA, 2023c]). Broadly, common features of the selected data are that they are available daily over CONUS, remotely accessible through an API, and produced on an ongoing basis. The suitable data products selected are summarized in **Table 1**.

2.2 Air quality

The air quality data products selected are the Goddard Earth Observing System Composition Forecast (GEOS-CF, https://gmao.gsfc.nasa.gov/weather_prediction/GEOS-CF/) and the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2, https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/). Both are global gridded products made publicly available by the NASA GSFC Global Modeling and Assimilation Office (GMAO) on an ongoing basis. GEOS-CF is a global three-dimensional model of atmospheric composition generated using the GEOS Earth System Model coupled with the GEOS-Chem chemical transport model, with initial meteorological conditions constrained by satellite observations [Keller et al., 2021]. Anthropogenic air pollutant emissions are obtained from pre-defined emission inventories [Crippa et al., 2018] and wildfire emissions are derived daily from satellite observations using the Quick Fire Emissions Dataset (QFED) [Darmenov and da Silva, 2015]. GEOS-CF produces daily 5-day global forecast at approximately 25 km x 25 km horizontal resolution of surface concentrations of criteria pollutants designated by the U.S. Environmental

Protection Agency (EPA) [USEPA, 2023]: particulate matter less than 2.5 microns in diameter (PM2.5), surface ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), and carbon monoxide (CO) [Keller et al., 2021; Knowland et al., 2022]. GEOS-CF contributes a novel forecasting capability to the suite of air quality information currently made available on the CDC Tracking Network. MERRA-2 is an atmospheric re-analysis product fusing measurements and model results to produce a physically-consistent estimate of the state of the atmosphere and aerosols for the period 1980-present at approximately 50 km x 50 km horizontal resolution with a latency of approximately 1 month behind real-time [Gelaro et al., 2017; Bosilovich et al., 2015; Randles et al., 2016, 2017; Buchard et al., 2016; 2017]. MERRA-2 uniquely contributes a long (+42 year) historical PM2.5 record and detailed information about the aerosol constituents (e.g., black carbon, organic carbon) of total PM2.5.

2.3 Flooding

The flood product selected is the Moderate Resolution Imaging Spectrometer (MODIS) near real-time (within 3-5 hours after observation) global flood product (MCDWD) [LANCE MCDWD, 2022]. MCDWD is a publicly available satellite-based data product that estimates the presence and surface extent of flood waters (i.e., standing water) over land (**Figure 1**). MCDWD is provided at ~250-meter resolution, with near-global daily coverage spanning 2011-present and is derived from the MODIS Surface Reflectance (MOD09) datasets from the Aqua and Terra satellites [NASA, 2023d, k; Slayback, 2023]. MCDWD is processed and made publicly available within 3 hours of collection by NASA's Land, Atmosphere Near real-time Capability for EOS (LANCE, https://lance.modaps.eosdis.nasa.gov/). Detailed descriptions of the MCDWD product, algorithms, product evaluation, data access, and planned improvements are available in Slayback

[2023] and online https://www.earthdata.nasa.gov/global-flood-product. Near real-time high-resolution (3-meter) flood maps derived from commercial PlanetScope imagery [Policelli, 2022] were considered but excluded because their application is presently more suited for limited areas of interest and routinely generating maps covering CONUS is not yet feasible.

3. Implementation & Data Systems Methodology

Through an interagency partnership, the goal is to define and establish routine processes for transforming and transferring NASA GSFC data to the CDC Tracking Program and its partners in an automated and timely fashion for dissemination via the CDC Tracking Network. Processes are developed in the CDC Enterprise Data Analytics and Visualization (EDAV) Platform, which is a cloud-based data management and processing ecosystem where users can store, transform, and analyze data. EDAV is built primarily using Microsoft Azure cloud services. Azure's implementation of Databricks, a web-based analytics platform, is used for scripting data ingestion from NASA APIs, further transformation of gridded raw data to geopolitical boundaries, and calculating relevant quantities (e.g., sum, mean, area) using Python and R. Completed data are routed to the Azure Data Lake, where most recent and archival versions of daily data runs are stored. Azure Data Factory is used to orchestrate the initialization of Databricks scripts in the appropriate order and the routing of data between different storage environments (**Figure 2**).

3.1 Air quality

The air quality data products are queried remotely, spatially transformed from model grid to U.S. counties or census tracts, depending on the dataset, and converted to measures tailored to environmental public health audiences. An automated daily query pulls GEOS-CF via Open-source Project for a Network Data Access Protocol (OPeNDAP), retrieving global 5-day forecasts for PM2.5, O3, SO2, NO2, and CO. Forecast results are extracted for the surface (model level 1) for CONUS and transformed from the GEOS-CF model grid to U.S. counties based on boundaries derived from Census TIGER 2000 [NOAA, 2023] (see **Section 6** for planned updates). Daily statistics (24-hour mean [PM2.5, SO2], 24-hour maximum [SO2, NO2], and maximum 8-hour rolling mean [CO, O3]) are calculated from hourly values for each county. An automated monthly query pulls MERRA-2 M2T1NXAER from the Goddard Earth Sciences Data and Information Services Center (GES DISC) [GMAO, 2015]. Surface (model level 1) aerosol constituents, SO2, and CO concentrations are extracted for CONUS. PM2.5 concentrations in micrograms per cubic meter are estimated by summing the aerosol constituents following [Buchard et al., 2017]

$$PM2.5 = PM2.5[DU] + PM2.5[SS] + PM2.5[OC] + PM2.5[BC] + (132.14/96.06)*PM2.5[SO4]$$

where DU is dust, SS is sea salt, OC is organic carbon, BC is black carbon, SO4 is sulfate, and the factor of 132.14/96.06 is applied to convert sulfate ion (molar mass of 96.06 g mol^{-1}) concentration output by MERRA-2 to ammonium sulfate (132.14 g mol^{-1}), assuming that sulfate is primarily present as neutralized ammonium sulfate. MERRA-2 SO2 surface concentrations are converted from kilograms per cubic meter (kg m^{-3}) to parts per billion (ppb) using the following equation:

$$SO2\ (ppb) = (T(k)^*\ SO2\ (kg\ m^{-3})^*8.314^*10^7) / (1.013^*64.066)$$

Coordinated Universal Times are converted to the local time of each county centroid, then daily statistics (24-hour mean [PM2.5 (and constituents), SO2], 24-hour maximum [SO2, NO2], and maximum 8-hour rolling mean [CO]) are calculated from hourly values for each county. The Air Quality Index (AQI) and AQI category (reflecting level of health concern) associated with each daily concentration for both GEOS-CF and MERRA-2 estimates are determined as specified in the U.S. EPA technical documentation [USEPA, 2018]. Concentrations higher than the AQI scale ("beyond the AQI") are classified as if they are in the highest AQI category [USEPA, 2018].

County-level daily AQI category estimates for GEOS-CF and MERRA-2 air pollutant estimates are mapped on the CDC Tracking Network's Data Explorer (**Figure 3**). Measures incorporating GEOS-CF data are housed within the "Forecasted Air Quality" indicator and those derived from MERRA-2 can be found in the "Current and Historical Air Quality" indicator, both of which are in the "Air Quality" content area. The "Forecasted Air Quality" indicator relies entirely on GEOS-CF data that updates each morning at approximately 7 am U.S. Eastern Standard Time (EST) and includes a 1-day hindcast and 4-day forecast (including current day). MERRA-2 is incorporated in a suite of daily air quality measures that also includes monitor data from the U.S. EPA's AirNow platform (https://www.airnow.gov/) [CDC, 2023b]. These measures include data from the previous several months, though only AirNow data is available for the present month due to the 1-month latency of MERRA-2. Users can view the underlying concentration used to derive the AQI category by hovering the cursor over a county of interest.

All AQI category and concentration data are also available in tabular format using the data download tool on the Data Explorer or via the CDC Tracking Network API.

3.2 Flooding

The MCDWD flood product is used to generate the "Total Area Flooded" and "Total Population Affected by Flooding" measures on the CDC Tracking Network's Data Explorer. Measures incorporating MCDWD data are housed within the "Current and Historical Flooding" indicators in the "Precipitation & Flooding" content area. The data are queried remotely, mosaicked, clipped, and spatially summarized from raster data to U.S. counties, U.S. territory county equivalents, and census tracts. A daily automated query retrieves MCDWD from the NASA LANCE system (https://lance.modaps.eosdis.nasa.gov/). The query is made at 3 am EST to retrieve the most-recent 2-day composite. The data are provided by LANCE in 10 degree x 10 degree tiles, with variables for two types of flood conditions: recurring flooding (available in 2024, see Section 6) and unusual flooding. The total area flooded is estimated by determining the proportion of the MCDWD tiles where flood has been identified relative to the total raster overlap with each jurisdiction. This proportion is then multiplied by the total area of the jurisdiction. Separate calculations are made to determine the quantity of recurring and unusually flooded land. To estimate the total number of people affected by flooding within each county and census tract, the sum of the population of all the affected census blocks within each county and tract is calculated. A block is considered affected if a flooded tile is identified anywhere within its boundary. Similarly, separate calculations are made to determine the total population affected by recurring and unusual flooding.

4. Applied Uses

Data integrated into the CDC Tracking Network are broadly intended to inform both research and early warning activities, including (1) describing temporal and spatial trends in disease and potential environmental exposures, (2) identifying populations most affected, (3) generating hypotheses about associations between health and environmental exposures, and (4) developing, guiding, and assessing environmental public health policies and interventions aimed at reducing or eliminating health outcomes associated with environmental factors. Numerous contextual data are available on the Data Explorer that can be viewed contemporaneously to help drive data-informed decision-making, such as biomonitoring data, disease burdens, community characteristics, and environmental justice indicators. Hypothetical use cases are provided below for illustrative purposes.

- **Target public health surveillance**: In the CDC Tracking Network's Data Explorer, the "Total Area Flooded" measure derived from MCDWD can provide situational awareness and identify where to target public health surveillance, community services, environmental sampling, or additional imagery acquisition (e.g., drone, commercial satellite). The maps can assist in identifying inundated areas, as well as dry areas where emergency response assets can be staged. The Data Explorer also has points of interest, such as nursing homes and day care centers, that can be viewed alongside a map of "Total Area Flooded" to better inform public health activities. Environmental justice measures, such as populations that are linguistically isolated, can also be layered on the Data Explorer map to help prioritize assistance.
- **Reduce personal exposure:** In summer 2023, a historic Canadian wildfire smoke event left millions of residents in the northeastern U.S. under air quality advisories [NASA,

2023e]. During a similar major air quality, an individual with asthma might consult the "Forecasted Air Quality" indicator derived from GEOS-CF on the CDC Tracking Network's Data Explorer alongside AirNow and other sources of information to inform the timing and confidence around decisions of when (or if) to perform strenuous exercise outdoors.

- **Examine trends in patient-centered outcomes:** A longitudinal study of wildfire smoke impacts on patient-centered outcomes for adult patients with asthma might examine the "Current and Historical Air Quality" indicators derived from MERRA-2 on the CDC Tracking Network's Data Explorer for insight on spatiotemporal trends in smoke exposure, cumulative exposures, timing and duration of smoke events, as well as statistical anomalies. In the Data Explorer, measures of PM2.5 and its constituents can be viewed alongside indicators of social vulnerability, occupation, and environmental justice indices to illuminate the interplay of wildfire smoke exposure and social determinants of health. The study results might then inform improved interventions such as refined public health messaging to populations who are at higher risk for asthma-related emergency department visits or hospitalizations during wildfire events.

5. Limitations

5.1 Air Quality

The GEOS-CF 5-day air quality forecast and MERRA-2 historical re-analysis provide regional estimates of air quality. Greater caution is warranted when interpreting results in geographic areas with strong pollution gradients (e.g., urban areas), mountainous terrain, locations near emission point sources, and locations on the coast. On average, GEOS-CF and MERRA-2 have greater skill in the eastern U.S. and in rural areas [Randles et al., 2016; Buchard et al., 2017]. MERRA-2 has less skill during winter and spring and in environments dominated by nitrate aerosol (e.g., intense agricultural areas) because nitrate is not explicitly accounted for in the GOCART aerosol mechanism [Randles et al., 2016]. MERRA-2 is constrained by aerosol optical depth (AOD), and there is inherent uncertainty in how aerosols are vertically distributed in the model. For GEOS-CF, forecast skill is generally greater for the nearer-term forecasts (i.e., Day 1, Day 2) [Keller et al., 2021]. For both GEOS-CF and MERRA-2, relative changes and trends are more robust than absolute changes. The air quality measures displayed in the Data Explorer represent outdoor ambient air pollution concentrations and do not directly reflect human exposures. The relationships between outdoor ambient concentrations, indoor air pollution, and individual exposures are active subjects of ongoing research, and likely vary depending upon pollutant, behavioral patterns, microenvironments, and building ventilation [Özkaynak et al., 2009; Patel et al., 2020; Singer et al., 2020]. The NASA air quality data products disseminated via the CDC Tracking Network are research-grade products and should not be used to assess National Ambient Air Quality Standards (NAAQS) compliance or to evaluate progress toward attaining compliance.

5.2 Flooding

MCDWD provides estimates of surface flood water extent at the time of satellite overpass between roughly 70N–70S. Greater caution is warranted in interpreting results for MODIS data under shadows caused by terrain and clouds, which can strongly resemble water in the spectral bands (red and near infra-red) used to generate the product [Slayback, 2023]. Flash floods have a low likelihood of being observed due to their rapid appearance and disappearance, unless this coincides with the twice-daily observation times of the Terra and Aqua satellites at approximately 10:30 am and 1:30 pm local solar time [Gosset et al., 2023]. Flooding in urban areas is also difficult for the 250-meter resolution MCDWD product to detect, due to the mismatch in sensor resolution to on-the-ground flooded areas. In addition to shadows, dark volcanic rock and recently melted snow can result in false positives, and vegetation cover can result in false negatives in MCDWD. The MODIS instruments aboard Terra and Aqua have flown in space for over 20 years and these missions have been scheduled for decommissioning in the 2025-2026 timeframe. MCDWD is transitioning to ingest data from the Visible Infrared Imaging Radiometer Suite (VIIRS) aboard the joint NASA/NOAA Suomi-NPP and NOAA-20 satellites.

6. Future Planned Work

Planned air quality-related updates include: improving the computational efficiency and flexibility of the Python workflow used to spatially aggregate MERRA-2 and GEOS-CF data; adding coverage over Alaska, Hawaii, Puerto Rico, and Virgin Islands; and updating the county and census boundary shapes to Census TIGER 2020. The update to Census TIGER 2020 boundaries is anticipated to mostly impact Alaska and Bedford, Virginia.

Planned flooding updates include: MCDWD recurring flood classification and exploring a second, complimentary flood product. The NASA MCDWD team plan to update the product to include a flood classification for “recurring flood” (water occurring where it has occurred in the past with some regularity, but is not permanent water) in early 2024. Maps on the CDC Tracking Network’s Data Explorer will be updated with new data when available. The HydroSAR flood product is being explored to potentially augment MCDWD daily maps. HydroSAR HyP3-

watermap is a 30-meter surface water extent product with a 12-day revisit period based on European Space Agency's (ESA) Sentinel-1 C-band synthetic aperture radar (SAR) [ASF, 2023]. The HydroSAR project focuses on cloud-based SAR data processing for rapid response and mapping of hydrological hazards. HydroSAR maps are retrieved by querying Alaska Satellite Facility's (ASF) Hybrid Pluggable Processing Pipeline (HyP3) service [Hogenson et al., 2020]. HydroSAR's revisit time is expected to be reduced to near-daily coverage with the planned launches of ESA's Sentinel-1C and NASA-Indian Space Research Organization (ISRO) Synthetic Aperture Radar (NISAR) in 2024. For public health surveillance and practice, the unique contributions of HydroSAR would be an order of magnitude increase in spatial resolution (MCDWD 250-meter vs. HydroSAR 30-meter), and SAR technology has all-weather and day-and-night imaging capabilities.

7. Conclusions

Through a partnership between the CDC Tracking Program and NASA GSFC that was established in 2022, critical environmental data products are delivered into the CDC Tracking Network's nationwide network of integrated health and environmental data in order to drive actions to improve community health. Near real-time monitoring and forecasting of environmental hazards have the potential to reduce harmful exposures by enabling early warning systems and other timely public health interventions. Currently, near real-time NASA data may not be readily accessible to public health practitioners, particularly those that need to react to public health emergencies, because the data are fragmented across multiple non-interoperable data systems and are not formatted in a manner consistent with typical public health uses. Developing data pipelines and processes to host these data in a usable and standardized format and disseminating these data via the CDC Tracking Network assists in making important information about environmental health hazards more readily accessible to public health practitioners, decision makers, and partners.

Based on CDC and partner priorities, NASA GEOS-CF daily 5-day air quality forecasts, MERRA-2 historical daily PM2.5 concentrations, and MCDWD daily near real-time maps of flooding are integrated into the CDC Tracking Network's publicly accessible Data Explorer (https://ephtracking.cdc.gov/DataExplorer/). These data products are mature, well-established resources produced on an ongoing basis and can be remotely accessed with automated queries to deliver timely CONUS-wide maps of environmental hazards. With the forthcoming decommissioning of the Aqua/Terra satellites, MCDWD will transition to ingesting VIIRS/Suomi NPP observations to ensure continuity. Augmenting MCDWD flood maps with other satellite-based flooding data products is being explored. The air quality and flooding data products newly integrated into the CDC Tracking Network are broadly intended to promote improve community health through data-informed decisions by informing trends, identifying potentially affected populations, and informing interventions that support better patient-centered outcomes.

Acknowledgements

Funding from CDC and GSFC Earth Sciences Division is gratefully acknowledged, as well as the SSAI-sponsorship of student interns. We acknowledge Michael Bosilovich and Diane Davies for technical insights into MERRA-2 data and LANCE, respectively. The NASA Goddard Modelling and Assimilation Office's (GMAO) research, development, and production of MERRA-2 is supported by the NASA Modeling, Analysis, and Prediction Program. HydroSAR is funded by the NASA Applied Sciences Program. The HydroSAR project is led by the University of Alaska Fairbanks and collaborates with interdisciplinary universities and NASA Centers, including NASA GSFC and Marshall Space Flight Center, ASF, and University of Alabama.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All data sources in this manuscript are free and publicly available. NASA GEOS-CF five-day forecasts of air quality pollutants are accessed via OpenDAP [Keller et al., 2021; GMAO, 2023]. NASA MERRA-2 historical aerosol concentrations are accessed via GES DISC [GMAO, 2015]. The NASA MCDWD daily near real-time flood product is accessed via LANCE [LANCE MCDWD, 2022]. Open-source software is used to query the data, spatially transform the data, and compute measures consumable to a public health audience (e.g., AQI). The code is written in R [R Foundation, 2022] and Python [Python Software Foundation, 2021]. The air quality and flooding measures are available as CONUS-wide maps and as downloadable data layers on the CDC Environmental Public Health Tracking Network Data Explorer [CDC, 2023a].

Author Contributions

- **Conceptualization:** H. Amos, N. Skaff, A. Werner, S. Schollaert Uz
- **Formal analysis:** F. Policelli, K. Patel, P. Abue, E. Macorps
- **Funding acquisition:** A. Werner, S. Schollaert Uz

- **Methodology:** N. Skaff, D. Slayback, M. Jo, C. Keller, V. Buchard, H. Amos
- **Supervision:** A. Werner, S. Schollaert Uz, H. Amos, N. Skaff

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or National Aeronautics and Space Administration.

References

Alaska Satellite Facility (ASF) (2023). HyP3 homepage, https://hyp3-docs.asf.alaska.edu/. Accessed: 1 May 2023.

Bosilovich, M. G., R. Lucchesi, and M. Suarez (2016). MERRA-2: File Specification. GMAO Office Note No. 9 (Version 1.1), 73 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes.

Buchard, V., Da Silva, A. M., Randles, C. A., Colarco, P., Ferrare, R., Hair, J., et al. (2016). Evaluation of the surface PM2.5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States. *Atmospheric Environment*, 125, 100–111, http://doi.org/10.1016/j.atmosenv.2015.11.004.

Buchard, V., C. A. Randles, A. da Silva, A. S. Darmenov, P. R. Colarco, R. C. Govindaraju, R. A. Ferrare, J. W. Hair, A. Beyersdorf, L. D. Ziemba, and H. Yu (2017). The MERRA-2 Aerosol

Reanalysis, 1980-onward, Part II: Evaluation and Case Studies. *Journal of Climate*, 30, 6851-6872, https://doi.org/10.1175/JCLI-D-16-0613.1.

CDC (2023a). CDC National Environmental Public Health Tracking Program website, https://www.cdc.gov/nceh/tracking/index.html. Accessed: 27 October 2023.

CDC (2023b). CDC National Environmental Public Health Tracking, Indicator: Historical Air Quality, https://ephtracking.cdc.gov/indicatorPages?selectedContentAreaAbbreviation=11&selectedIndicatorId=203. Accessed: 27 October 2023.

Census (2021). US Census Bureau TIGER/Line Shapefiles, released 2 February 2021, https://www.census.gov/geographies/mapping-files/2020/geo/tiger-line-file.html.

Crippa, M., D. Guizzardi, M. Muntean, E. Schaaf, F. Dentener, J.A. van Aardenne (2018). Gridded emissions of air pollutants for the period 1970-2012 within EDGAR v4.3.2. *Earth System Science Data*, 10(4): 1987–2013. https://doi.org/10.5194/essd-10-1987-2018.

Darmenov, A.S. and A. da Silva (2015). The Quick Fire Emissions Dataset (QFED) - Documentation of versions 2.1, 2.2 and 2.4. Technical report series on Global Modeling and Data Assimilation, NASA//TM-2015-104606 (Vol. 38, p. 212). Available online: https://gmao.gsfc.nasa.gov/pubs/docs/Darmenov796.pdf.

Duncan, B.N., C.A. Malings, E. Knowland, D.C. Anderson, A.I. Prados, C.A. Keller, K.R. Cromar, S. Pawson, and H. Ensz (2021). Augmenting the Standard Operating Procedures of Health and Air Quality Stakeholders with NASA Resources, *GeoHealth*, 5(9), https://doi.org/10.1029/2021GH000451.

Gelaro, R. et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *Journal of Climate*, 30(14), 5419-5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Global Modeling and Assimilation Office (GMAO) (2015). MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly, Time-averaged, Single-Level, Assimilation, Aerosol Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), [Dataset] DOI: 10.5067/KLICLTZ8EM9D (Accessed on 9 Nov 2023)

Global Modeling and Assimilation Office (GMAO) (2023). GEOS Composition Forecasts (GEOS-CF) Data Access webpage, https://gmao.gsfc.nasa.gov/weather_prediction/GEOS-CF/data_access/ (Accessed on 9 Nov 2023)

Gosset, M., P.A. Dibi-Anoh, G. Schumann, R, Hostache, A, Paris, E.-P. Zahiri, M. Kacou, and L. Gal (2023). Hydrometeorological Extreme Events in Africa: The Role of Satellite Observations for Monitoring Pluvial and Fluvial Flood Risk, *Surveys in Geophysics*, 44, 197-223, https://doi.org/10.1007/s10712-022-09749-6

Hogenson, K., Kristenson, H., Kennedy, J., Johnston, A., Rine, J., Logan, T., Zhu, J., Williams, F., Herrmann, J., Smale, J., and F. Meyer (2020). Hybrid Pluggable Processing Pipeline (HyP3): A cloud-native infrastructure for generic processing of SAR data [Software]. Zenodo. https://doi.org/10.5281/zenodo.4646138

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob. D. J., and Pawson, S. (2021). Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0. *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002413, [Dataset] DOI: https://doi.org/10.1029/2020MS002413 (Accessed on 9 Nov 2023)

Knowland, K.E., C.A. Keller, and R. Lucchesi (2022). File Specification for GEOS-CF Products. GMAO Office Note No. 17 (Version 1.3), 54 pp, available from https://gmao.gsfc.nasa.gov/pubs/office_notes.php

LANCE MCDWD (2022). NRT MODIS/Aqua+Terra Global Flood Product MCDWD_L3_NRT distributed from NASA LANCE. Available on-line: https://www.earthdata.nasa.gov/global-flood-product, [Dataset] DOI: 10.5067/MODIS/MCDWD_L3_NRT.061 (Accessed on 9 Nov 2023)

Liu, A.Y., J.M. Trtanj, E.K. Lipp, and J.M. Balbus (2021). Toward an integrated system of climate change and human health indicators: a conceptual framework, *Climatic Change*, 166(40), https://doi.org/10.1007/s10584-021-03125-w.

NASA (2023a). NASA Goddard Space Flight Center Sciences and Exploration Directory Earth Sciences Division Code 610 website, https://science.gsfc.nasa.gov/earth/. Accessed: 8 March 2023.

NASA (2023b). NASA EarthData Floods Data Pathfinder, https://www.earthdata.nasa.gov/learn/pathfinders/disasters/floods-data-pathfinder. Accessed: 8 March 2023.

NASA (2023c). NASA EarthData Data Processing Levels, https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-information-policy/data-levels. Accessed: 8 March 2023.

NASA (2023d). NRT MODIS/Aqua+Terra Global Flood Product MCDWD_L3_NRT distributed from NASA LANCE. Available online, https://www.earthdata.nasa.gov/global-flood-product. DOI: 10.5067/MODIS/MCDWD_L3_NRT.061.

NASA (2023e). Smoke Smothers the Northeast, *Earth Observatory*, https://earthobservatory.nasa.gov/images/151433/smoke-smothers-the-northeast. Accessed: 8 June 2023.

NOAA (2023). National Weather Service U.S. Counties shapefile, https://www.weather.gov/gis/Counties. Accessed: 15 June 2021.

Özkaynak, H., B. Glenn, J.R. Qualters, H. Strosnider, M.A. McGeehin, and H. Zenick (2009). Summary and findings of the EPA and CDC symposium on air pollution exposure and health. *Journal of Exposure Science & Environmental Epidemiology*, 19(1):19–29.

Patel, S., S. Sankhyan, E.K. Boedicker, P.F. DeCarlo, D.K. Farmer, A.H. Goldstein, E.F. Katz, W.W. Nazaroff, Y. Tian, J. Vanhanen, and M.E. Vance (2020) Indoor particulate matter during HOMEChem: Concentrations, size distributions, and exposures. *Environmental Science & Technology*, 54(12):7107–16.

Policelli, F. (2022), NASA GSFC Flood Mapping System (In beta testing). Presented at the American Geophysical Union Fall Meeting 2022, Chicago, Illinois. Available online: https://ntrs.nasa.gov/citations/20220015132

Python Software Foundation (2021). 3 May 2021 Release (Version 3.9.5) [Software], https://docs.python.org/release/3.9.5/

R Foundation (2022).10 March 2022 Release (Version 4.1.3) [Software], https://cran.r-project.org/src/base/R-4/

Randles, C. A., A. da Silva, V. Buchard, P. R. Colarco, A. S. Darmenov, R. C. Govindaraju, A. Smirnov, R. A. Ferrare, J. W. Hair, Y.Shinozuka, and C. Flynn (2017). The MERRA-2 Aerosol Reanalysis, 1980-onward, Part I: System Description and Data Assimilation Evaluation. *Journal of Climate*, 30(17), 6823-6850, https://doi.org/10.1175/JCLI-D-16-0609.1.

Randles, C.A., A. M. da Silva, V. Buchard, A. Darmenov, P. R. Colarco, V. Aquila, H. Bian, E. P. Nowottnick, X. Pan, A. Smirnov, H. Yu, and R. Govindaraju (2016). The MERRA-2 Aerosol Assimilation, Technical Report Series on Global Modeling and Data Assimilation, Volume 45, NASA/TM–2016-104606/Vol. 45, available from https://gmao.gsfc.nasa.gov/pubs/docs/Randles887.pdf.

Slayback, D. (2023). MODIS NRT Global Flood Product User Guide Revision C. Available from https://www.earthdata.nasa.gov/s3fs-public/2023-01/MCDWD_UserGuide_RevC.pdf.

Singer, B.C., W.R. Chan, Y.-S. Kim, F.J. Offermann, and I.S. Walker (2020). Indoor Air Quality in California Homes with Code-required Mechanical Ventilation, *Indoor Air*, 30(5): 885-899, https://doi.org/10.1111/ina.12676.

US EPA (2018). Technical Assistance Document for the Reporting of Daily Air Quality - the Air Quality Index (AQI). EP 454/B-18-007, September 2018. Available from https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf.

US EPA (2023). US EPA Criteria Air Pollutants website, https://www.epa.gov/criteria-air-pollutants. Accessed: 8 March 2023.

USGCRP (2016). The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment. Crimmins, A., J. Balbus, J.L. Gamble, C.B. Beard, J.E. Bell, D. Dodgen, R.J. Eisen, N. Fann, M.D. Hawkins, S.C. Herring, L. Jantarasami, D.M. Mills, S. Saha, M.C. Sarofim, J. Trtanj, and L. Ziska, Eds. U.S. Global Change Research Program, Washington, DC, 312 pp. http://dx.doi.org/10.7930/J0R49NQX

Watts, N., M. Amann, N. Arnell et al. (2021). The 2020 report of *The Lancet* Countdown on health and climate change: responding to converging crises. *The Lancet*, 397(10269), 129-170, https://doi.org/10.1016/S0140-6736(20)32290-X.

WMO (2023). World Metrological Organization, Early Warnings for All, https://public.wmo.int/en/earlywarningsforall. Accessed: 27 October 2023.

Figures and Tables

Corresponding author: Helen Amos, helen.m.amos@nasa.gov

Table 1. NASA Goddard research-grade products for air quality and flooding implemented into the CDC Tracking Program's workflow.

Data product	**Resolution**	**Strengths and limitations for public health uses**
GEOS-CF air quality forecast	Daily 5-day forecast 25 km x 25 km grid, global	+ daily forecast of criteria pollutants - coarse resolution
MERRA-2 aerosols	Daily, 1980-present 50 km x 50 km, global	+ 40-year record, daily resolution - coarse resolution, 1-month lag
MODIS NRT global daily flood product (MCDWD)	Daily, 2022-present 250 m, global	+ low latency, daily coverage - cannot see through clouds

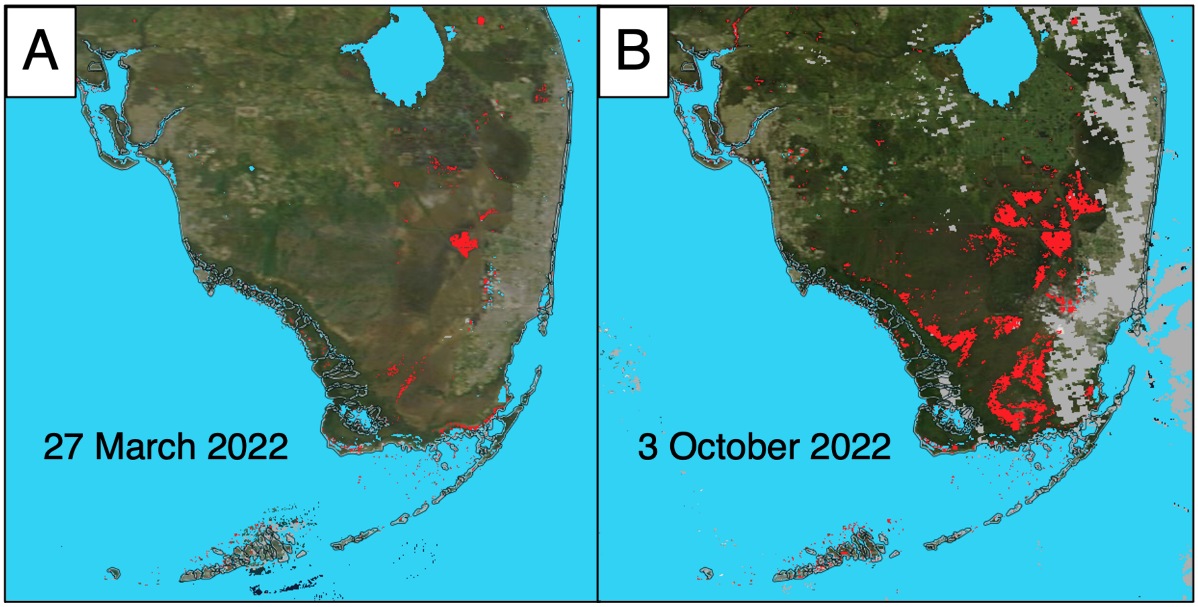

Figure 1. Flooding before (A) and after (B) Hurricane Ian as mapped by the NASA MODIS near real-time global flood product (MCDWD). Flooded pixels are red, gray pixels are clouds. Ian made landfall on 28 Sep 2022 near Fort Meyers, Florida as a Category 4 hurricane. (A) 27 March 2022 is shown to illustrate a representative baseline of conditions before hurricane season. (B) Not all the flood water detected on 3 October 2022 is from Hurricane Ian. Flood water can persist for weeks or longer.

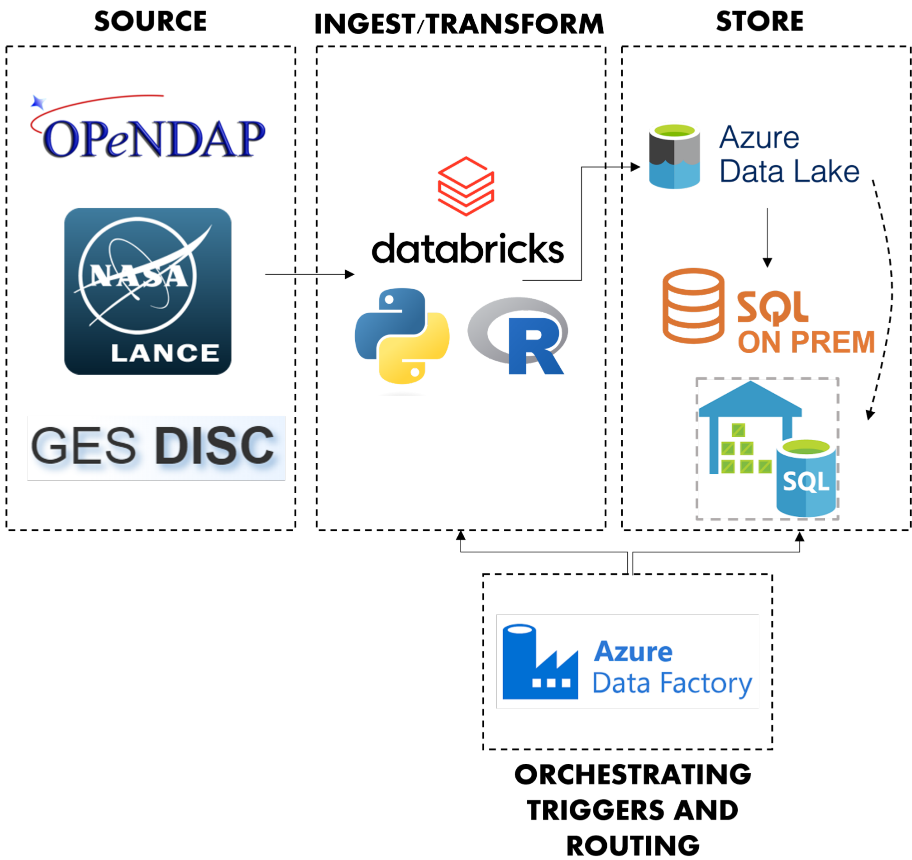

Figure 2: Schematic of the CDC Tracking Network's extract, transform, and load (ETL) workflow.

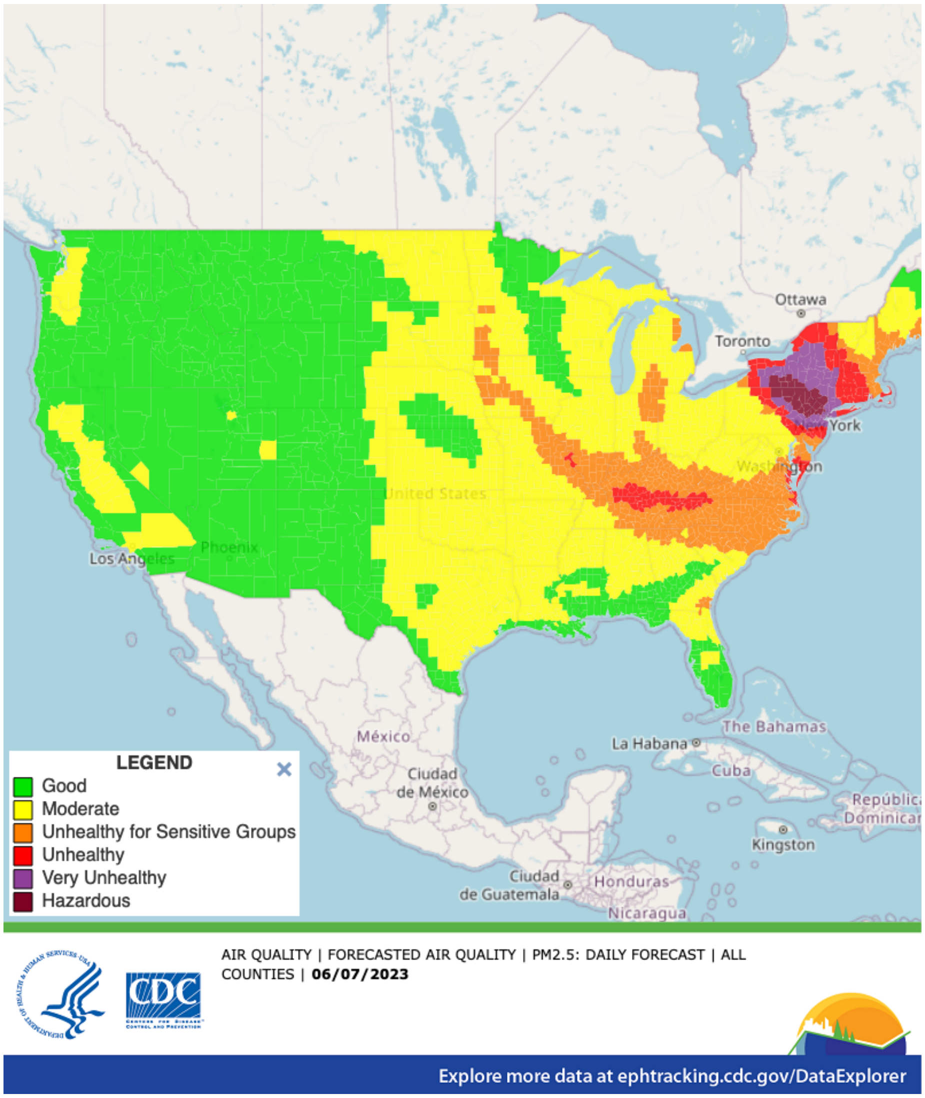

Figure 3. Example of a map of GEOS-CF forecasted PM2.5 as seen on the CDC Tracking Network's publicly available Data Explorer. On this day, 6 June 2023, heavy smoke from Canadian wildfires impacted major population centers in the eastern U.S.