Data in Brief 58 (2025) 111208

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Expediated modeling of burn events results (EMBER): A screening-level dataset of 2023 ozone fire impacts in the US

Heather Simon[a,*], James Beidler[b], Kirk R. Baker[b], Barron H. Henderson[a,b], Loren Fox[a], Chris Misenis[a], Patrick Campbell[c,d], Jeff Vukovich[a], Norm Possiel[a], Alison Eyth[a]

[a] *Office of Air and Radiation, US Environmental Protection Agency, 109 TW Alexander Dr, PO Box 12055, RTP, NC 27711, USA*

[b] *Office of Research and Development, US Environmental Protection Agency, 109 TW Alexander Dr, PO Box 12055, RTP, NC 27711, USA*

[c] *Cooperative Institute for Satellite Earth System Studies, Center for Satellite and Earth Science Research (CSER), George Mason University, Fairfax, VA, USA*

[d] *Air Resources Laboratory, National Oceanic and Atmospheric Administration, College Park, MD, USA*

ARTICLE INFO

Article history:
Received 10 September 2024
Revised 18 November 2024
Accepted 2 December 2024
Available online 6 December 2024

Dataset link: Expediated Modeling of Burn Events Results (EMBER) Data Files (Original data)

Keywords:
CMAQ
Source attribution
Air pollution
Zero-out
Photochemical modeling

ABSTRACT

The Expedited Modeling of Burn Events Results (EMBER) dataset consists of 36-km grid-spacing Community Multiscale Air Quality (CMAQ) photochemical modeling for the summer of 2023. For emissions, these simulations utilized representative monthly and day-of-week anthropogenic emissions from a recent year and preliminary day-specific 2023 fire emissions derived using BlueSky pipeline. The base model run simulated ozone concentrations across the contiguous US during Apr 11-Sep 29, 2023. Two zero-out model runs simulated ozone levels that would have occurred in the US (1) in the absence of fire emissions (“Zero Fires”) and (2) in the absence of only Canadian wildfire emissions (“Zero Canadian Fires”). Fire impacts on ozone were then estimated as the difference between ozone simulated in the base EMBER run compared to the ozone simulated in each of the zero out model runs. EMBER is presented as a screening level

---

[*] Corresponding author.
E-mail address: Simon.Heather@epa.gov (H. Simon).

https://doi.org/10.1016/j.dib.2024.111208

dataset due to the emissions limitations and the 36-km grid-spacing used in these simulations.

Specifications Table

Subject	Environmental Science: Air Pollution
Specific subject area	*2023 Fire Impacts on Ground-Level Ozone Concretions in the US*
Type of data	Table Figures Binary (NetCDF file format) ASCII csv
Data collection	Fire emissions were derived using the US Forest Service BlueSky modeling framework. Ozone data were produced using the US Environmental Protection Agency's Community Multiscale Air Quality (CMAQ) model. Fire impacts on ozone were determined using zero-out modeling.
Data source location	Institution: U.S. Environmental Protection Agency City/Town/Region: Research Triangle Park NC Country: USA
Data accessibility	Repository name: Zenodo... Data identification number: 10.5281/zenodo.13737753 Direct URL to data: 10.5281/zenodo.13737754
Related research article	None

1. Value of the Data

- These data provide photochemical model-based estimates of fire impacts on US ozone concentrations during the summer of 2023.
- The model-based estimates allow for exploration of the types of conditions that lead to substantial ozone production from fire emissions and comparison of those conditions to ones where fire plumes do not form large amounts of ozone.
- Estimates from photochemical modeling can be compared to estimates of ozone fire impacts produced by statistical models being reported in the literature [1]. Comparing results from independent methods will allow the community to better quantify uncertainties and limitations of each technique and to identify conditions under which each method produces the most credible results.

2. Background

Wildfires burned a record number of acres in Canada during the spring and summer of 2023. Smoke plumes from these fires impacted locations across the United States and led to an unusual number of fire impacted pollution episodes in some areas. Fires emit a range of air pollution components including nitrogen oxides (NO_X = NO + NO2) and volatile organic compounds (VOCs). Both NO_X and VOCs are precursors to ozone which is harmful to human health. States and local agencies may exclude certain high-pollution days from regulatory determinations if they can demonstrate, among other points, that a clear causal relationship exists between ozone and an event that is not reasonably controllable or preventable, such as a wildfire [2]. In 2023, fire emissions from Canada mixed with other natural and anthropogenic US emissions making it challenging to quantify the portion of ozone that originated from the fire plumes without sophisticated modeling tools. Using modeling tools to support such determinations requires time and resources. Here we present a screening-level modeling dataset as a proof-of-concept to show how existing datasets can be leveraged to produce rapid estimates of fire impacts on ozone. The dataset can be explored to further understand the controlling factors leading to ozone formation from fire emissions during this extreme fire season.

3. Data Description

Data documented in this article have been archived through Zenodo [3]. The following data files are described in this section: Anthropogenic and fire emissions files, model runscripts and ancillary files, gridded model output files, and figures and tables of model results. A full description of the model set-up including information on how to obtain meteorology and initial and boundary conditions files is provided in the Experimental Design, Materials and Methods section.

3.1. Anthropogenic and fire emissions

Emissions files are provided in a zipped tar file: EMBER.EMIS.tar.gz. When unzipped there is a folder for each emissions sector. The individual files are NetCDF version 3 formatted files using I/O API data structures (IOAPI) with meta data described at https://www.cmascenter.org/ioapi/.

Anthropogenic emissions for 2018 [4] were supplied for two gridded sectors and six point-source sectors listed below.

Emissions from the following sources are included in the provided files: US, Canadian and Mexican anthropogenic emissions from sources within the modeling domain, US prescribed and wildfire emissions, Canadian wildfire emissions. Mexican wildfire emissions were not included. As described in Experimental Design, Materials and Methods section, the biogenic emissions were computed inline in the model simulations and are also not included in the emissions input files.

Anthropogenic emissions from US, Canadian, and Mexican sources within the modeling domain were provided for 3 representative days in each month from April through September as listed in Table 1.

- Gridded emissions sectors
  - merged_nobeis_norwc: these files included all gridded emissions sources except for biogenic and residential wood combustion emissions.
  - rwc: these files included gridded residential wood combustion emissions.
- Point source sectors
  - ptnonipm: these files included all point source emissions not included in one of the other five point source categories.
  - pt_oilgas: these files included point source emissions from oil and gas exploration and production activities.

Table 1
Days from the 2018 emissions files used as the representative day for each day of the week and the month in 2023.

Month	Monday–Friday	Saturday	Sunday
April	April 3, 2018	April 7, 2018	April 8, 2018
May	May 7, 2018	May 12, 2018	May 13, 2018
June	June 5, 2018	June 9, 2018	June 10, 2018
July	July 9, 2018	July 14, 2018	July 15, 2018
August	August 7, 2018	August 11, 2018	August 12, 2018
September	September 11, 2018	September 15, 2018	September 16, 2018

Table 2
Monthly 2023 Canada wildland fire emissions (short tons).

Month	CO (tons)	CO2 (tons)	NOx (tons)	PM2.5 (tons)	VOC (tons)
Jan	174,640	2943,797	2509	24,299	54,063
Feb	26,430	466,961	426	3716	8673
Mar	33,191	597,026	564	4765	11,244
Apr	29,440	569,269	601	4228	10,448
May	4530,476	76,361,970	65,021	607,398	1346,389
Jun	8572,781	138,754,014	112,729	1134,010	2442,094
Jul	6592,839	112,494,448	101,312	883,722	2046,350
Aug	5765,409	98,807,492	89,586	780,091	1809,509
Sep	3871,245	65,528,376	56,429	527,038	1177,193
Oct	327,395	5690,758	5044	45,253	102,992
Total	**29,923,847**	**502,214,113**	**434,222**	**4014,519**	**9008,956**

  - othpt: these files included all point source emissions located in Mexico and Canada within the 36US3 domain.
  - cmv_c3_36: these files included all point source emissions from large ocean-going c3 class ships within the 36US3 domain.
  - cmv_c1c2_36: these files include all point source emissions from smaller c1 and c2 class ships within the 36US3 domain.
  - ptegu: these files include all point source emissions from electric generating units.

2023 fire emissions were supplied for 3 point source sectors for each modeled day:

- ptfire_canada: these files include all Canadian fires located withing the 36US3 domain.
- ptfire-wild: these files included all US wildfires.
- ptfire-rx: these files included all US prescribed fires.

Point source sectors include two types of files:

- inline_mole: files beginning with this prefix provide the emissions rates for each emitted pollutant at each point source emissions location. There is a separate “inline_mole” file for each day with emissions.
- stack_groups: files beginning with this prefix provide metadata for each point source including location information (i.e. latitude, longitude, location within the model grid) and stack parameters used to determine plume rise within the model (height, diameter, exit velocity, exit temperature). There is a single “stack_groups” file covering all days for each anthropogenic emissions sector. For fire sectors there is a separate “stack_groups” file for each day.

Fig. 1 and Tables 2 and 3 provide summaries of 2023 fire emissions for $PM_{2.5}$, CO, CO_2, NO_X, and VOC.

3.2. Gridded model output files

Three IOAPI gridded post-processed model output files are provided:

- HR2DAY_CMAQv54_cb6r5_ae7_aq.36US3.35.basecase2023_2023_O3_PM25_11Apr2023to29 Sep2023.nc: this file represents model results for the base EMBER model simulation that included all emissions.
- HR2DAY_CMAQv54_cb6r5_ae7_aq.36US3.35.basecase2023_nofires_2023_O3_PM25_11Apr202 3to29Sep2023.nc: this file represents model results from a “zero-out” simulation that was identical to the base EMBER simulation except that no point source fire emissions were included in the simulation (“Zero Fires”).

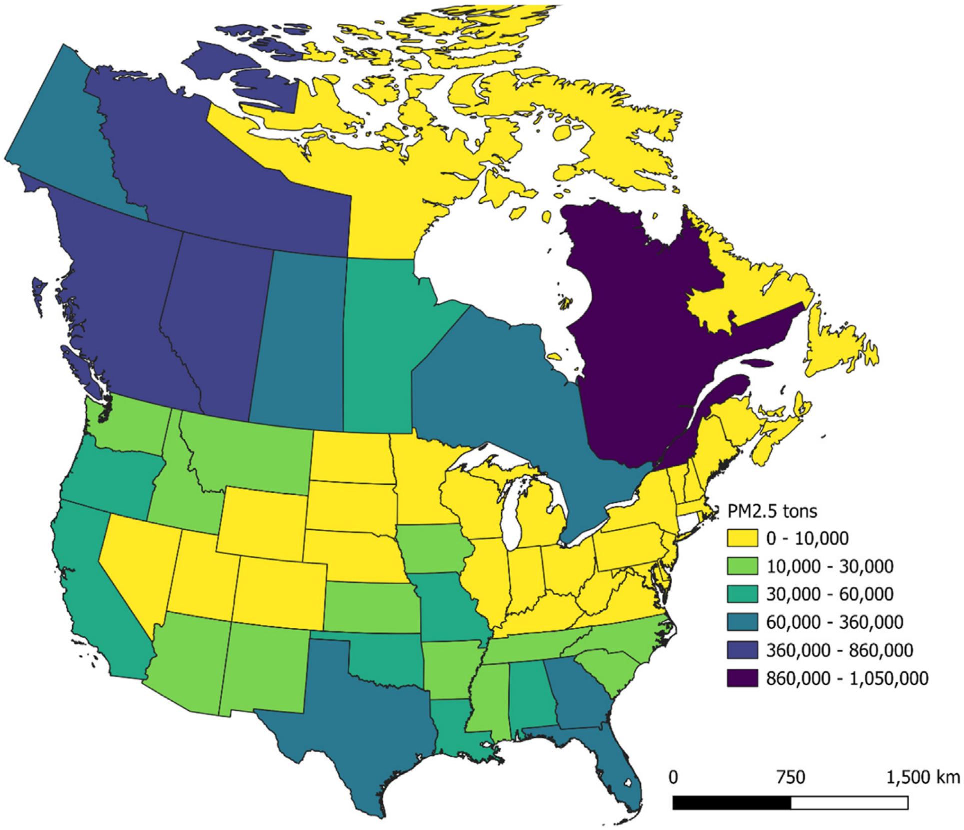


Fig. 1. State and province total January – October 2023 wildland fire emissions of primary $PM_{2.5}$ (short tons).

- HR2DAY_CMAQv54_cb6r5_ae7_aq.36US3.35.basecase2023_noCAfires_2023_O3_PM25_11Apr2023to29Sep2023.nc: this file represents model results form a “zero-out” simulation that was identical to the base EMBER simulation except that it did not include point source emissions from the Canadian wildfire sector (fire emissions from US wild and prescribed fires were included in this simulation) (“Zero Canadian Fires”).

Each file includes gridded 2-dimensional (Layer 1) post-processed daily model outputs for 172 timesteps (one timestep for each day from April 11, 2023-Sep 29, 2023). The following variables are included in each file:

- O3_MDA8: 8-hr daily maximum ozone concentration (ppb) calculated in accordance with Appendix U to 40 CFR Part 50.
- ATOTIJ_AVG: 24-hour average of fine particulate matter (accumulation and Aiken modes) ($\mu g/m^3$)

3.3. Table and figures of results

Model data were post-processed to create several types of figures and tables. These figures and tables are included in the Zenodo data archive [3] in the file called: EMBER_figures_table.zip with the following contents:

Table 3
Monthly 2023 CONUS wildland fire emissions.

Month	CO (tons)	CO2 (tons)	NOx (tons)	PM2.5 (tons)	VOC (tons)
Jan	244,127	5681,144	8997	54,146	65,202
Feb	409,920	9128,348	14,228	87,566	101,336
Mar	694,207	15,087,298	23,022	152,435	162,633
Apr	476,004	11,146,661	17,767	103,474	124,714
May	267,886	6077,854	9331	55,523	78,595
Jun	211,376	4288,445	6054	42,738	57,528
Jul	225,447	4624,932	6482	40,114	69,601
Aug	538,032	10,258,753	13,056	95,992	172,104
Sep	287,245	5841,789	7941	53,329	88,950
Oct	506,505	9670,550	12,324	92,955	146,199
Total	**3860,750**	**81,805,773**	**119,202**	**778,272**	**1066,862**

- ozone_mda8_paired_merge.v2.31may2024.forQlik.xlsx
- folder: Maps of ozone fire impacts
- folder: Ozone monitor & model overlay maps
- folder: PM monitor & model overlay maps
- folder: monitor-level MDA8 ozone timeseries

Each file type is described in more detail below. In addition to the Zenodo archive, these figures and tables are accessible via a Qlik web application which allows the user to easily query and subset data and browse subsets of relevant figures. The Qlik application is accessible at: https://awsedap.epa.gov/public/single/?appid=ef2c5326-6e19-4c1f-94f6-bd5a63cf240f&sheet=fGzBmv&theme=horizon&opt=ctxmenu,currsel&identity=preview.

Ozone data table: The ozone data table contains observed and modeled ozone data paired in space and time. The data table includes the following fields:

- **SiteId** - AQS monitor ID in format SSCCCMMMM where SS = state FIPS, CCC = county FIPS, MMMM = monitor ID.
- **POCode** - Parameter Occurrence Code. Used to distinguish different instruments that measure the same parameter at the same site.
- **Date** - date range of Apr 11-Sep 29, 2023.
- **State** - state name.
- **County** - county name.
- **cbsa_name** - name of the core based statistical area.
- **csa_name** - name of the combined statistical area.
- **2015 Ozone Nonattainment Area** - 2015 ozone nonattainment area name, if applicable.
- **Monitored MDA8 ozone (ppb)** - monitored MDA8 ozone.
- **Modeled MDA8 ozone (ppb): Base Simulation** - modeled MDA8 ozone from base EMBER simulation for the 36 km grid cell in which the monitor is located.
- **Modeled MDA8 Ozone Impacts from US and Canada Fires (ppb)** - EMBER predictions for the 36 km grid cell in which the monitor is located of MDA8 ozone attributed to US and Canadian wild and prescribed fire emissions. Calculated as [Modeled MDA8 ozone (ppb): Base Simulation] – [Modeled MDA8 ozone (ppb): Zero Fires Simulation].
- **Modeled MDA8 Ozone Impacts from Canadian Wildfires (ppb)** - EMBER predictions for the 36 km grid cell in which the monitor is located of MDA8 ozone attributed to Canadian wildfire emissions. Calculated as [Modeled MDA8 ozone (ppb): Base Simulation] – [Modeled MDA8 ozone (ppb): Zero Canada Fires Simulation].
- **Modeled MDA8 ozone (ppb): Zero Fires Simulation** - modeled MDA8 ozone for the 36 km grid cell in which the monitor is located from EMBER simulation without any emissions from US and Canadian wild and prescribed fires.

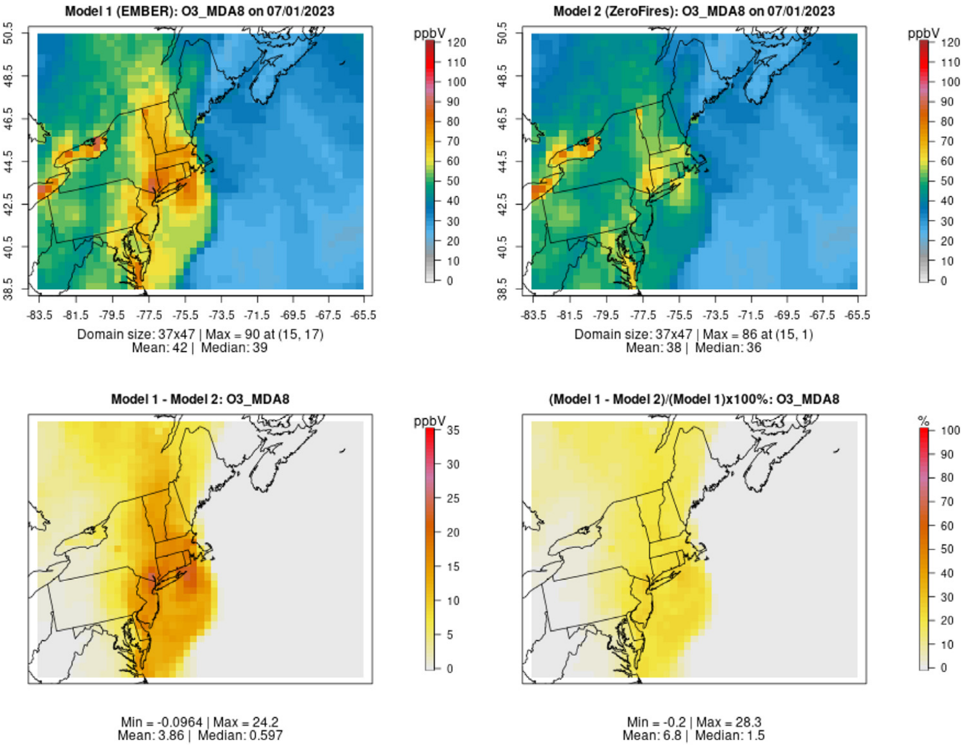


Fig. 2. Ozone fire impacts modeled in EMBER on July 1, 2023 in the Northeast region. MDA8 ozone concentration (ppb) modeled in the base EMBER simulation (top left); MDA8 ozone (ppb) modeled in the EMBER simulation without and US or Canadian fire emissions (ZeroFires) (top right); Absolute impact of fire emissions on MDA8 ozone concentration (ppb) calculated as EMBER – ZeroFires (bottom left); Relative impact of fire emissions on MDA8 ozone concentration (%) calculated as 100 × (EMBER – ZeroFires)/EMBER (bottom right).

- **Modeled MDA8 ozone (ppb): Zero Canada Wildfires Simulation** - modeled MDA8 ozone for the 36 km grid cell in which the monitor is located from EMBER simulation without any emissions from Canadian wildfires.
- **Network** - ozone monitoring network name.
- **monitor_type** - type of organization operating the monitoring site (i.e. EPA vs state or local government agencies (SLAMS) vs industry etc.).
- **Latitude** - location of ozone monitor given in decimal degrees.
- **Longitude** - location of ozone monitor given in decimal degrees

Maps of ozone fire impacts: US regional maps were created for every day of the simulation period showing fire impacts on 8-hr daily maximum (MDA8) ozone concentrations (172 days × 9 regions = 1548 figures). NOAA climate regions [5] were used to define regions for plotting: Northeast, Northern Rockies and Plains, Northwest, Ohio River Vally, South, Southeast, Southwest, Upper Midwest, and West. Fig. 2 provides an example map of ozone fire impacts of July 1, 2023 in the Northeast region.

Ozone monitor & model overlay maps: US regional maps were created for every day of the simulation period showing monitored MDA8 ozone concentrations (ppb) overlaid on top of gridded modeled MDA8 ozone concentrations (ppb) (172 days × 9 regions = 1548 figures). Fig. 3 provides an example of this type of plot for July 1, 2023 in the Northeast region.

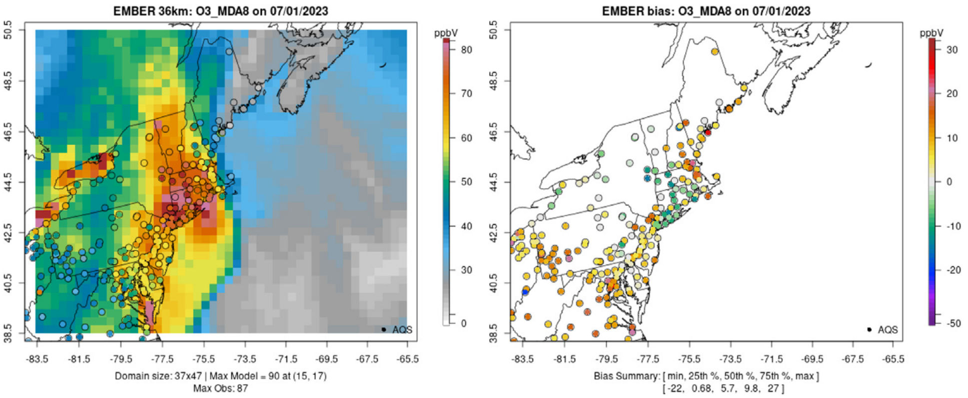


Fig. 3. Measured MDA8 ozone concentrations (ppb) shown at monitoring locations (circles) on top of gridded modeled MDA8 ozone concentrations (ppb) for July 1, 2023 (left). Model bias MDA8 ozone concentration (ppb) at monitor locations calculated as model – observations for July 1, 2023 (right).

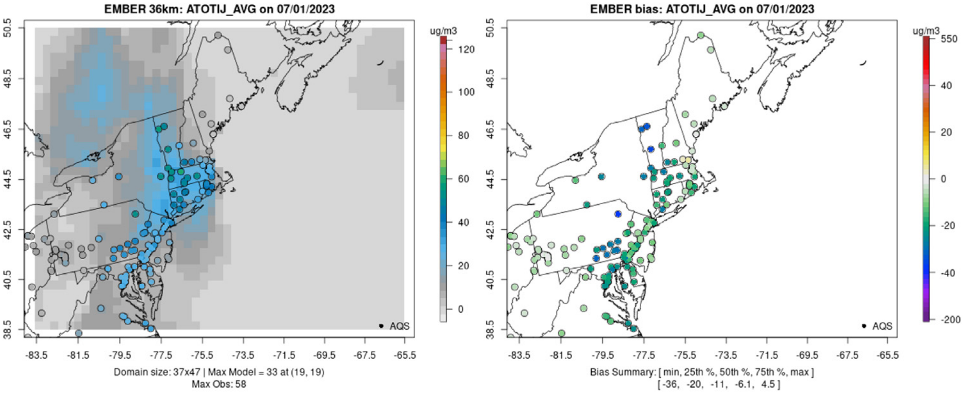


Fig. 4. Measured 24-h daily average $PM_{2.5}$ concentrations (μg/m^3) shown at monitoring locations (circles) on top of gridded modeled 24-h daily average $PM_{2.5}$ concentrations (μg/m^3) for July 1, 2023 (left). Model bias for 24-h daily average $PM_{2.5}$ concentrations (μg/m^3) at monitor locations calculated as model – observations for July 1, 2023 (right).

PM monitor & model overlay maps: US regional maps were created for every day of the simulation period showing monitored 24-hr daily average $PM_{2.5}$ concentrations (μg/m^3) overlaid on top of gridded modeled 24-hr daily average $PM_{2.5}$ concentrations (μg/m^3) (172 days × 9 regions = 1548 figures). Fig. 4 provides an example of this type of plot for July 1, 2023 in the Northeast region.

Monitor-level MDA8 ozone timeseries: Ozone timeseries plots were created at every ozone monitoring location (1210 figures). Timeseries show modeled and observed MDA8 ozone concentrations at the monitor location as well as modeled fire impacts as shown in Fig. 5 for an ozone monitor located in Adams County, PA.

3.4. Model runscripts and ancillary files

Model runscripts and ancillary files used in the model simulation are provided in the EMBER_inputs_and_scripts.zip file. Contents of this zip file include:

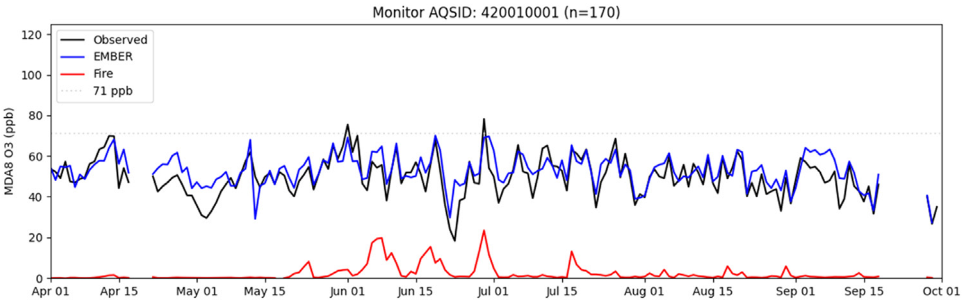


Fig. 5. Observed (black) and modeled (blue) MDA8 ozone concentration (ppb) at Adams Count, PA monitor ID: 420,010,001 from Apr-Sep 2023. Modeled fire impacts on MDA8 ozone (ppb) shown in red. Any points above the dotted line at 71 ppb indicate MDA8 ozone values that exceed the 2015 US EPA National Ambient Air Quality Standard (NAAQS) level.

- Concat.sh: bash script used to adjust timesteps for the meteorological inputs files to match the 0Z to 0Z CMAQ run duration for each day.
- GRIDDESC: ascii file that includes grid specifications for the modeling domain.
- LNTG_AllParms_36US3.ioapi: IOAPI formatted file for calculating inline lightning NO emissions [6].
- merged.dates.2023.to.2018_clean.txt: ascii file that provides the mapping of representative 2018 emissions days to 2023 calendar days.
- runscript_EMBER_base: CMAQ model runscript.
- ssmask.36US3.ncf: ocean mask file used for the inline calculation of sea salt emissions.
- folder: inline biogenic emissions inputs – folder contains files necessary to for calculation of inline biogenic emissions using BEIS4 model as described here: https://github.com/USEPA/CMAQ/blob/main/DOCS/Users_Guide/CMAQ_UG_ch06_model_configuration_options.md#6.9.2_Online_Emission
  - beis4_norm_emis.36US3.ncf
  - gspro_biogenics_19nov2019_nf_v10.txt
- folder: namelist files – folder contains all necessary namelist files used at model runtime as described at: https://github.com/USEPA/CMAQ/blob/main/DOCS/Users_Guide/CMAQ_UG_ch04_model_inputs.md
  - AE_cb6r5_ae7_aq.nml
  - CMAQ_Control_DESID.nml
  - CMAQ_Control_DESID_cb6r5_ae7_aq_NVOA.nml
  - GC_cb6r5_ae7_aq.nml
  - NR_cb6r5_ae7_aq.nml
  - Species_Table_TR_0.nml
- folder: photolysis input files – folder contains all necessary input files for CMAQ photolysis calculations as described here: https://github.com/USEPA/CMAQ/blob/main/DOCS/Users_Guide/CMAQ_UG_ch06_model_configuration_options.md#6.10.3_Photolysis
  - CSQY_DATA.F
  - OMI_1979_to_2023aug.dat
  - PHOT_OPTICS.dat

4. Experimental Design, Materials and Methods

4.1. Modeling domain and episode

The Expedited Modeling of Burn Event Results (EMBER) simulations were run for the periods of April 1 through September 30, 2023. However, the first 10 days of April were used as

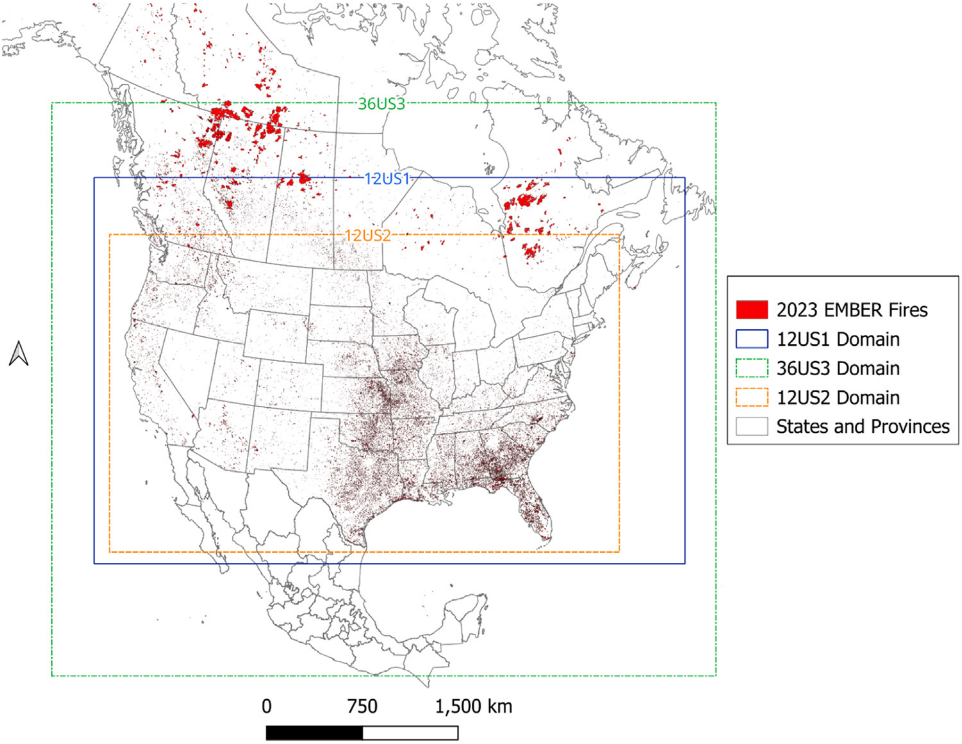
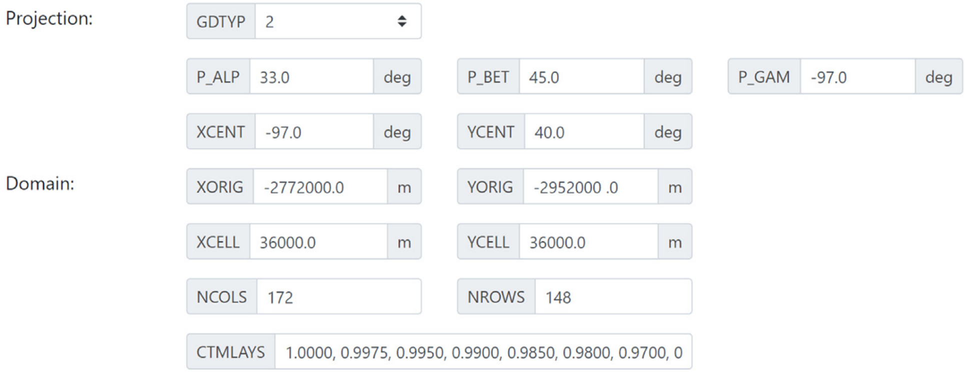


Fig. 6. 2023 wildland fire activity and model domain boundaries. 12US1 and 12US2 domains were not used in this dataset but are shown for reference as they are commonly applied modeling domains over the United States.

model spin-up and are not included in the results datasets. Additionally, because the calculation of MDA8 ozone values requires some hours from the following morning and since the model simulation ended at 11:59PM Greenwich Mean Time (GMT) on Sep 30th, there were not a sufficient number of hours to calculate valid MDA8 ozone values on Sep 30th. Consequently, MDA8 ozone outputs are only provided through Sep 29, 2023. The April 11 through September 29 time period was sufficient to cover most of the wildfire activity impacts on ozone for days on which there were monitored violations of the ozone National Ambient Air Quality Standards during 2023.

EMBER utilized EPA's 36US3 modeling domain denoted by the green box in Fig. 6. This domain was chosen for several reasons. First, the domain extends far enough North to cover the location of most of the Canadian wildfires that occurred in 2023. Second, the 36-km grid-spacing provides for rapid runtime and reasonable output file sizes which make quick turnaround of these simulations feasible. The map in Fig. 6 also shows the location of burn perimeters for 2023 along with the boundaries of several 12-km model domains that have been used previously by the United States Environmental Protection Agency (US EPA) for research and regulatory applications.

This grid domain is based on a Lambert conformal conic (LCC) projection centered at (−97,40) with true latitudes of 33 and 45° north. In LCC meters, the 36US3 domain uses 36 km grid-spacing that extends in the east-west direction from 2,952,000.000 m to 3,204,000.000 m in 172 columns and in the north-south direction from −2,772,000.000 m to 2,520,000.000 m in 148 rows. The modeling domain included 35 vertical layers extending up to 50 millibar (mb) using a hybrid terrain-following and constant-pressure structure.



Projection:
GDTYP 2
P_ALP 33.0 deg
P_BET 45.0 deg
P_GAM -97.0 deg
XCENT -97.0 deg
YCENT 40.0 deg

Domain:
XORIG -2772000.0 m
YORIG -2952000 .0 m
XCELL 36000.0 m
YCELL 36000.0 m
NCOLS 172
NROWS 148
CTMLAYS 1.0000, 0.9975, 0.9950, 0.9900, 0.9850, 0.9800, 0.9700, 0

Fig. 7. Screenshot of NACC cloud application options to generate MCIP files.

4.2. Modeling inputs

In this section, we describe model inputs that were used for the EMBER simulations. These include meteorological data, initial and lateral boundary conditions, anthropogenic emissions, fire emissions and other miscellaneous model input files.

4.3. Meteorological inputs

EMBER used meteorology provided by NOAA's Global Forecast System (GFS FV3; https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php last accessed 2024–07–10) model, which was run at approximately 13-km grid-spacing with 127 vertical levels. We use the NOAA-EPA Atmosphere-Chemistry Coupler (NACC) application to create CMAQ model-ready meteorological fields on the EMBER 36-km grid. These data can be generated via downloading and compiling the NACC code (GitHub - noaa-oar-arl/NACC: NOAA-EPA Atmosphere-Chemistry Coupler [7]) or by using the cloud-based application: (https://nacc.arl.noaa.gov/nacc/setup [8]). For our application, we used the cloud version that had access to the required hourly GFS file via Amazon Web Services. If using the offline version, the necessary NACC inputs from GFS are available upon request from NOAA. Using the cloud-based system, the user can generate CMAQ-ready meteorological inputs outputs for the EMBER grid by utilizing the options outlined in Fig. 7 and Table 4.

The other options should default to the correct values. The only values that should be modified outside of the values listed above will be the dates of the period to be modeled.

Daily meteorological outputs from NACC run from 12 Greenwich Mean Time (GMT) to 12 GMT the next day. The CMAQ model simulations were conducted to run from 0 GMT to 0 GMT. To address this mismatch in time steps, EMBER concatenated 2 days' worth of meteorology files for each daily meteorological input file and extracted the relevant hours in a temporary file using the concatmet.sh bash script available in the Zenodo data archive [3]. The concatmet.sh requires runtime inputs of the day 1 filename, day 2 file name, output file name and the date.

4.4. Initial conditions and lateral boundary conditions

The initial and lateral boundary conditions were derived from NASA's Goddard Earth Observing System composition forecast (GEOS-CF) global forecast model [9]. The GEOS-CF model

Table 4
vertical layer structure of CMAQ simulations.

CMAQ Layer	sigma
35	0.000
34	0.050
33	0.100
32	0.150
31	0.200
30	0.250
29	0.300
28	0.350
27	0.400
26	0.450
25	0.500
24	0.550
23	0.600
22	0.650
21	0.700
20	0.740
19	0.770
18	0.800
17	0.820
16	0.840
15	0.860
14	0.880
13	0.900
12	0.910
11	0.920
10	0.930
9	0.940
8	0.950
7	0.960
6	0.970
5	0.980
4	0.985
3	0.990
2	0.995
1	0.9975
Surface	1.000

runs daily for one “hindcast” day and five “forecast” days. We use the hindcast day, which includes meteorological assimilation or “analysis.” Currently, the hindcast does not include composition assimilation beyond what goes into the meteorological system (e.g., ozone column). We use the “analysis” day as lateral boundary conditions by extracting the nearest GEOS-CF cell to each boundary cell around the CMAQ domain perimeter. Data is extracted from the hourly average meteorological (met), chemical (chm), and extra GEOS-Chem Chemistry (xgc) files through GMAO’s OpenDAP server. The composition data is then translated from GEOS-Chem speciation to CMAQ cb6r3/ae7.

At present, the processor will download data every third hour and interpolate to create hourly input files for CMAQ. The entire process was done using the geoscf2bc package available at: https://github.com/barronh/geoscf2bc/, but in the future, the capability has been ported to the aqmbc package available at: https://github.com/barronh/aqmbc/. The aqmbc package uses the same general process for GEOS-CF and other models, so it uses a more generalized approach. Both packages require Python3.6 or above and can be installed using the Python Installation Package (pip). The commands to install and run a day using geoscf2bc are provided below.

Code Listing: geoscf2bc Installation on Linux

```
python -m pip install -qq git+https://github.com/barronh/geoscf2bc.git
```

Code Listing: geoscf2bc 1-day script

```
from geoscf2bc.drivers import default
# This example makes its own GRIDDESC. Normally, you use your own.
with open('GRIDDESC', 'w') as gf:
  gf.write("""' '
'LamCon_40N_97W'
  2 33.000 45.000 -97.000 -97.000 40.000
' '
'36US3'
'LamCon_40N_97W'  -2952000.0 -2772000.0 36000.0 36000.0 172 148 1
' '""")
outpaths = default(
    GDNAM='36US3', gdpath='GRIDDESC',
    SDATE='2023-04-01T00', EDATE='2024-04-02T00'
)
```

4.5. Fire emissions

Since the primary aim of this modeling was to predict impacts from wildland fires, it was determined that timely day-specific fire emissions estimates were crucial for this assessment. To support the modeling, EPA developed a set of draft US and Canada 2023 wildland fire emissions. This modeling did not include any fire emissions for the portions of Mexico within the 36-km domain. The applied wildland fire emissions estimate methods were consistent with the process described in the technical support documentation of the 2020 National Emissions Inventory (NEI) [10]. The development of emissions estimates during the fire season meant that retrospective datasets typically used to develop wildland fire activity were not yet available. US wildfire perimeter information was obtained from the Wildland Fire Interagency Geospatial Services (WFIGS) current interagency fire perimeters dataset, while Canada wildfire perimeter information was obtained from the National Burned Area Composite (NBAC) geospatial dataset. The wildfire perimeter data were reconciled with Hazard Mapping System (HMS) satellite detections using a modified python version of SmartFire 2. Emissions estimates were calculated from the fused wildland fire activity using the series of fuel loading, consumption, and emissions calculations modules in BlueSky pipeline. Canadian fuel loading was obtained by mapping Canada's Fuel Behavior Prediction (FBP) fuel beds to Fuel Characteristic Classification System (FCCS) loadings. All wildland fire emissions used in the 2023 modeling are considered preliminary and are expected to be replaced with emissions developed for the 2023 NEI.

The 2023 US and Canada wildland fire emissions area available at as part of the Zenodo data archive [3].

4.6. U.S., Canadian, and Mexican anthropogenic emissions

For anthropogenic emissions, EPA used month-specific representative weekday emissions for each Monday through Friday, along with Saturday and Sunday emissions by month for 2018 from the "2018v2" inventory [4], the most recently developed regulatory emissions platform available (https://www.epa.gov/air-emissions-modeling/2018v2-emissions-modeling-platform). We were not able to utilize day-specific 2023 anthropogenic emissions that would account for submonthly changes such as increased power plant emissions on high energy demand days. However, using representative days from a recent inventory allowed us to simulate a typical chemical environment in which the daily fire emissions could react to form ozone. Table 1 provides the list of representative days used for the day of the week and the month.

Agricultural fire emissions from 2018v2 were not included in these simulations to avoid any potential double counting with the 2023 fire emissions described above. Agricultural fire emissions generally account for a small percentage of total emissions for NO_X and VOC. In 2020 NEI [10], the most recent NEI data available from the EPA (https://www.epa.gov/air-emissions-inventories/2020-national-emissions-inventory-nei-data), agricultural fires accounted for 0.45 % of national NO_X emissions and 0.23 % of national VOC emissions.

4.7. Other model inputs

Additional model inputs used for EMBER simulations are listed below. These input files are available in the Zenodo data archive [3].

- LTNGPARMS: file used for calculating online lightning NO emissions [6].
- GRIDDESC file: provides specifications for the 36US3 model grid.
- OCEAN file: identifies the fractional coverage in each model grid cell allocated to open ocean for calculation of sea salt emissions.
- Files for calculating inline photolysis.
  - OMI: Ozone Monitoring Instrument column data used to determine how much light is absorbed by ozone above the model domain.
  - CSQY_DATA: contains the cross sections and quantum yields of photolysis rates used by the chemical mechanism.
  - OPTICS_DATA: describes the optical properties of clouds, aerosols, and the earth's surface.
- Namelist files
  - gc_matrix_nml: Namelist look-up tables for gas-phase species used to define the parameters of different model species during the execution of the CMAQ programs.
  - ae_matrix_nml: Namelist look-up tables for aerosol species used to define the parameters of different model species during the execution of the CMAQ programs.
  - nr_matrix_nml: Namelist look-up tables for non-reactive species used to define the parameters of different model species during the execution of the CMAQ programs.
  - tr_matrix_nml: Namelist look-up tables for tracer species used to define the parameters of different model species during the execution of the CMAQ programs.
  - DESID_CTRL_NML: DESID control namelist allowing using to customized rules for reading and emissions.
  - DESID_CHEM_CTRL_NML: links chemical variables on the emissions streams to CMAQ species, and offers scaling capabilities, geographic specicficity with the regions functionalities, size distribution customization, and the use of chemical families.
- Inline biogenic emissions files
  - GSPRO: speciation profiles file that contains the factors that are used to separate aggregated inventory pollutant emissions totals into emissions of chemical mechanism model species.
  - BEIS_NORM_EMIS: file with gridded normalized emissions calculated with both summer and winter emissions factors.

4.8. Model simulation configuration

EMBER model simulations were run with CMAQv5.4 [11] on 128 processors. The runscript is available at in the Zenodo data archive [3]. Key model configuration options used are provided below:

- Treatment of atmospheric transport and vertical flux
  - M3DRY dry deposition scheme without bi-directions ammonia flux.
  - WRF Pleim-Xu land-surface model.
  - Gravitational settling turned on.
  - enhanced ozone deposition over seawater turned off.
- Inline emissions calculations
  - BEIS4 inline biogenic emissions calculations.
  - Windblown dust emissions turned off.
  - Inline lightning NO emissions calculated based on statistical relationships with the simulated convective rainfall rate and scaled using World Wide Lightning Location Network (WWLLN) data.
- Chemistry and aerosol treatment
  - CB6r5 chemical mechanism.
  - Aero 7 aerosol module run with non-volatile primary organic aerosol and without potential combustion secondary organic aerosol formation.
  - Surface HONO chemistry turned on.

4.9. Post-processing of model outputs

EMBER model outputs were processed using multiple publicly available tools.

1. COMBINE: The COMBINE tool was used to convert raw ozone outputs in ppmv to units of ppbv and to calculate total $PM_{2.5}$ concentrations as the sum of all PM species. COMBINE is available at https://github.com/USEPA/CMAQ/tree/main/POST/combine.
2. HR2DAY: The HR2DAY tool was used to calculate MDA8 ozone concentrations and 24-h $PM_{2.5}$ concentrations from hourly data. HR2DAY is available at https://github.com/USEPA/CMAQ/tree/main/POST/hr2day.
3. Sitecmp and sitecmp_dailyO3: The sitecompare and sitecmp_dailyO3 programs were used to pair gridded model outputs with monitor data in space in time. These programs are available at https://github.com/USEPA/CMAQ/tree/main/POST/sitecmp and https://github.com/USEPA/CMAQ/tree/main/POST/sitecmp_dailyo3. Ambient monitoring data used to run these programs were pulled from EPA's Air Quality System (AQS) on May 23, 2024.

Limitations

There are several limitations to the EMBER modelling dataset. First the 36-km grid-spacing, while fine enough to simulate regional processes, may not fully capture local scale meteorological phenomena such as land-sea breezes and transport in complex terrain nor may it fully capture emissions and pollution gradients in urban areas or near emissions sources. Next, EPA has not yet completed the 2023 National Emissions Inventory, so EMBER instead utilized anthropogenic emissions that were developed for a recent year (2018). While these emissions provide a reasonable approximation of pollutant emissions in the proximity of 2023, they will not capture any emissions trends that have occurred in the 5 years between 2018 and 2023, nor will they capture day-specific emissions patterns that are driven by meteorological conditions in 2023 such as increased power usage during heat waves. Emissions estimates for wildland fires were limited by the quality of the data readily available. Wildland fire satellite and perimeter data were obtained from near real time sources that may not be representative of the final area and locations burned. Because of these limitations, we recommend EMBER be viewed as a screening dataset.

Model performance can be evaluated by comparing model estimates to measured pollutant concentrations. For this analysis we compared our ozone results against the range of bias and

Table 5
April 11-September 29, 2023, EMBER MDA8 ozone model performance statistics calculated as described in Simon et al. [12].

Region*	Normalized Mean Bias (%)	Normalized Mean Error (%)	Correlation
Northeast	12.9	17.2	0.76
Upper Midwest	2.9	14.2	0.77
Ohio River Valley	8.0	13.8	0.71
Southeast	15.2	17.6	0.79
Northern Rockies and Plains	1.4	10.6	0.62
South	4.4	14.1	0.78
Northwest	6.6	14.8	0.67
West	3.0	15.0	0.73
Southwest	−7.4	12.2	0.54

error values reported in the literature for recent state-of-the science model simulations [12,13]. Table 5 provides model performance statistics for MDA8 ozone concentrations against ozone monitors in the nine NOAA climate regions. These performance statistics are in the range of those reported for modeling studies by Simon et al. [12] and Emery et al. [13]. In addition, for the purpose of evaluating the ability of EMBER for capturing ozone impacts from fires, spatial and temporal patterns of known fire plumes were compared to model predictions of both ozone and $PM_{2.5}$. This comparison showed that the model generally captured the spatial extent, timing and magnitude of enhanced $PM_{2.5}$ and ozone concentrations measured by ground-level monitors in locations of known fire plumes. Figs. 3 and 4 demonstrate this for the July 1 plume in the Northeast US and further examples are available in the full set of figures provided with this dataset. It is recommended that users of this dataset supplement the regional performance statistics and general observations about simulated plume timings and locations with a more detailed comparison of ozone and $PM_{2.5}$ concentrations to measurement data at specific times and locations of interest for their specific application.

Ethics Statement

This research meets the ethical requirements for publication in Data in Brief. This work does not involve studies with animals and humans, or data collected from social media platforms.

Credit Author Statement

Heather Simon: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing, Visualization, Project Administration. **James Beidler:** Methodology, Validation, Investigation, Data Curation, Writing, Visualization. **Kirk Baker:** Conceptualization, Methodology, Validation, Investigation, Writing. **Barron Henderson:** Conceptualization, Methodology, Software, Validation, Investigation, Data Curation, Writing, Visualization. **Loren Fox:** Software, Writing. **Chris Misenis:** Investigation, Data Curation, Writing. **Patrick Campbell:** Methodology, Software, Investigation, Data Curation, Writing. **Jeff Vukovich:** Validation, Investigation, Writing. **Norm Possiel:** Conceptualization, Methodology, Validation, Writing. **Alison Eyth:** Methodology, Validation, Writing.

Data Availability

Expediated Modeling of Burn Events Results (EMBER) Data Files (Original data) (zenodo).

Acknowledgments

The authors would like to acknowledge David Mintz and Ben Wells for helping to process AQS data for ozone and PM2.5. The authors would like to thank Wyat Appel, Kristen Foley and Christian Hogrefe for helpful assistance with scripts for running CMAQ post-processing tools. The authors would like to acknowledge Tyler Fox, Beth Palma, Gabrielle Deabler, Gobeail McKinley and Vera Kornlyak for helpful conversations about the utility of model-based estimates of fire impacts on ozone in 2023.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

The views expressed in this paper are those of the authors and do not necessarily represent the views or policies of the US Environmental Protection Agency.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Lee, D.A. Jaffe, Wildfire impacts on O3 in the continental United States using PM2.5 and a generalized additive model (2018–2023)„ Environ. Sci. Technol. 58 (33) (2024) 14764–14774, doi:10.1021/acs.est.4c05870.
[2] Treatment of Data Influenced by Exceptional Events, 40 C.F.R. **§** 50.14 (2016)
[3] H. Simon, J. Beidler, K. Baker, B. Henderson, L. Fox, C. Misenis, P. Campbell, J. Vukovich, N. Possiel, A. Eyth, Expediated modeling of burn events results (EMBER) Data Files, Zenodo (2024), doi:10.5281/zenodo.13737754.
[4] United States Environmental Protection Agency. Technical Support Document (TSD): preparation of Emissions Inventories for the 2018v2 North American Emissions Modeling Platform. U.S. Environmental Protection Agency Office of Air Quality Planning and Standards, Research Triangle Park, NC, 2023; EPA-454/B-23-003. Available at: https://www.epa.gov/air-emissions-modeling/2018v2-emissions-modeling-platform-technical-support-document
[5] T.R. Karl, W.J. Koss, Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895–1983, Histor. Climatol. (1984) 38 Series 4-3pp.
[6] D. Kang, C. Hogrefe, G. Sarwar, J.D. East, J.M. Madden, R. Mathur, B.H. Henderson, Assessing the Impact of Lightning NOx Emissions in CMAQ Using Lightning Flash Data from WWLLN over the Contiguous United States, Atmosphere (Basel) 13 (2022) 1248, doi:10.3390/atmos13081248.
[7] P.C. Campbell, Y. Tang, P Lee, B Baker, D Tong, R Saylor, A Stein, J Huang, H.-C Huang, E Strobach, J McQueen, L Pan, I Stajner, J Sims, J Tirado-Delgado, Y Jung, F Yang, T.L Spero, R.C Gilliam, Development and evaluation of an advanced National air quality forecasting capability using the NOAA Global Forecast System version 16, Geosci. Model Dev. 15 (2022) 3281–3313, doi:10.5194/gmd-15-3281-2022.
[8] P.C. Campbell, W. Jiang, Z. Moon, S. Zinn, Y. Tang, NOAA's Global forecast system data in the cloud for community air quality modeling, Atmosphere (Basel) 14 (2023) 1110, doi:10.3390/atmos14071110.
[9] C.A. Keller, K.E Knowland, B.N Duncan, J Liu, D.C Anderson, S Das, R.A Lucchesi, E.W Lundgren, J.M Nicely, E Nielsen, L.E Ott, E Saunders, S.A Strode, P.A Wales, D.J Jacob, S. Pawson, Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0, J. Adv. Model. Earth Syst. 13 (2021) e2020MS002413, doi:10.1029/2020MS002413.
[10] United States Environmental Protection Agency. Technical Support Document (TSD) Preparation of Emissions Inventories for the 2020 North American Emissions Modeling Platform. 2023, U.S. Environmental Protection Agency Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/B-23-004. Available at: https://www.epa.gov/air-emissions-modeling/2020-emissions-modeling-platform-technical-support-document
[11] United States Environmental Protection Agency. CMAQ (Version 5.4) [Software]. 2022, Available from 10.5281/zenodo.7218076.
[12] H. Simon, K.R. Baker, S. Phillips, Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012, Atmos. Environ. 61 (2012) 124–139, doi:10.1016/j.atmosenv.2012.07.012.
[13] C. Emery, A. Liu, A.G.M. Russell, M.T. Odman, G. Yarwood, N. Kumar, Recommendations on statistics and benchmarks to assess photochemical model performance, J Air Waste Manag. Assoc 67 (5) (2017) 582–598, doi:10.1080/10962247.2016.1265027.