

Supplementary Information

NEIVA v.1.0:

Next-generation Emissions InVentory expansion of Akagi et al.

Samiha Binte Shahid¹, Forrest G. Lacey^{2,5}, Christine Wiedinmyer³, Robert J. Yokelson⁴, Kelley C. Barsanti^{1,5}

¹Department of Chemical & Environmental Engineering, Center for Environmental Research & Technology, University of California-Riverside, Riverside, 92521, United States

²Research Applications Laboratory, U. S. National Science Foundation National Center for Atmospheric Research, Boulder, 80301, United States

³Cooperative Institute for Research in Environmental Sciences, University of Colorado-Boulder, Boulder, CO, 80309, United States

⁴Department of Chemistry, University of Montana, Missoula, 59812, United States

⁵Atmospheric Chemistry Observations and Modeling Laboratory, U. S. National Science Foundation National Center for Atmospheric Research, Boulder, 80301, United States

Correspondence to: Kelley C. Barsanti (barsanti@ucar.edu)

Table of Tables

Table S1: Overview of the databases and datasets that comprise NEIVA, the backend database used in the Python script files for processing data and generating datasets, and querying features and functions, with reference to relevant sections and tables in the SI.	3
Table S2: List of tables comprising the legacy database (ldb). Abbreviations were assigned in this work and used in subsequent tables to denote fire or fuel type.	4
Table S3: Column headers and descriptions for tables in the legacy database (ldb).	5
Table S4: Types of unique id with examples.	6
Table S5: List of tables comprising the raw database (rdb). Abbreviations refer to one of the 14 fuel or fire types as defined in Table S2. For publications that include multiple fuel or fire types, no abbreviation for the fuel or fire type is given. Each table includes data from a single manuscript.	7
Table S6: Mapping of individual fuels from FLAME-4 and FIREX laboratory studies to fire type.	9
Table S7: Domestic biomass burning categories.	9
Table S8: Tables comprising the primary database (pdb).	11
Table S9: Average EFs based on Koss et al. (2018) and % C correction factor applied to account for actual % C reported by Selimovic et al. (2018).	14
Table S10: Columns of the NEIVA integrated dataset.	14
Table S11: A subset of the integrated EF dataset that represents groups of compounds with equivalent ids but different names.	15
Table 12: The subset of grouped compounds from Table S11 after non-unique groups of compounds were combined as described in text.	16
Table S13: The size of the integrated EF dataset through the integration steps.	17
Table S14: Reference publications used in the emission factor adjustment.	18
Table S15: Descriptions of NMOC_g datasets selected for fractional contribution assignment.	23
Table S16: Column headers and rows in the recommended EF dataset.	24
Table S17: Number and list of model surrogates used to represent the individual gaseous NMOCs.	24
Table S18: Mapping MOZART-T1 species to SAPRC-18 model species.	25
Table S19: Mapping GEOS-Chem model species to SAPRC-07 model species.	25
Table 20: The compounds with unassigned GEOSChem species.	26
Table S21: Comparison between the summation of EF of assigned and unassigned model species of GEOS-Chem.	27
Table S22: The description and units of property variables.	27
Table S23: Description tables comprises the backend database.	36
Table S24: Description of functions that displays information.	38
Table S25: Description of functions used to query EF data and properties of NMOC_g.	39

Overview of the NEIVA Databases, Datasets, and Querying Functions and Features

Table S1: Overview of the databases and datasets that comprise NEIVA, the backend database used in the Python script files for processing data and generating datasets, and querying features and functions, with reference to relevant sections and tables in the SI.			
Data Storage Name	Description	SI Section(s)	SI Table(s)
Legacy database (ldb)	The Akagi et al. (2011) supplemental data, including 2014 and 2015 updates, are stored as tables in this repository. There are 14 tables, one for each fire or fuel category.	S1. Legacy and Raw Databases	Table S2, Table S3
Raw database (rdb)	Data from selected publications (2015 or later) are stored as tables in this repository. There are 30 tables, one for each of the publications added since Akagi et al. (2011)	S1. Legacy and Raw Databases	Table S5
Primary database (pdb)	Data from the legacy and raw database tables were reformatted to achieve a consistent structure and combined with some additional data processing as described in the manuscript and SI. The resultant 44 tables are stored in this repository.	S2. Primary Database	Table S8
NEIVA output database (odb)	Integrated EF dataset: EF data aggregated in the primary database were merged and stored in this single dataset.	S3. Integrated EF Dataset	Table S10-Table S13
	Processed EF dataset: Several data processing steps were performed prior to generating averages for the recommended EF dataset and the resultant EFs are stored in this single dataset.	S4. Processed EF Dataset	Table S14
	Recommended EF dataset: The calculated averages of all EFs for each of the 14 fire and fuel types are stored in this single dataset.	S5. Recommended EF Dataset	Table S17
	Property_Surrogate dataset: Chemical and physical property data, as well as surrogate model species assignments, for each of the gaseous organic compounds in these datasets are stored in this single dataset.	S6. Chemical Property and Mechanism Surrogate Dataset	Table S18-Table S22
Backend database	Tables that are used in the Python scripts for generating datasets and querying data are stored in this database.	S8. Backend Database	Table S23
n/a	Descriptions of querying functions and features enabled by Python scripts.	S9. Querying Functions and Features	Table S24

S1. Legacy and Raw Databases

Data in the legacy and raw databases were extracted from the referenced publications and checked for duplicate values using a Python script file. Each of the supplemental tables in Akagi et al. (2011) were stored in NEIVA as a separate table as listed in Table S2. Additional details on the supplemental tables, including definitions of each fire type, can be found in Akagi et al. (2011). While most of the EF data were directly imported from Akagi et al. (2011) to the legacy database (ldb) there were two exceptions:

- The ‘akagi11_tropical_forest’ table included a single EF_HONO measurement made on a fire that had higher than average NO_x emissions. Since there were multiple NO_x measurements available, Akagi et al. (2011) used the single measured HONO/NO_x ratio times the average EF_NO_x to estimate the average EF_HONO. In the NEIVA ldb, this average EF_HONO replaced the single EF_HONO from Akagi et al. (2011).
- The ‘akagi11_temperate_forest’ table included nephelometer-based prescribed fire PM_{2.5} EFs from Burling et al. (2011). These were not retained in the ldb, and were replaced with new PM₁ EFs for the same fires based on AMS data from May et al. (2014) in the rdb.

The NEIVA legacy database retained only the information from Akagi et al. (2011) that is listed in Table S3 below; the footnotes and equations in the Akagi et al. (2011) supplemental tables were not retained.

In general, each row in the legacy and raw databases represents a compound or formula, and the attributes of that compound or formula are represented in the columns. In some of the tables, the first few rows are used to provide descriptive information about the measurements, such as modified combustion efficiency and analytical method. The columns in the raw database are the same as listed in Table S3 for the legacy database, with the exception of the UNC and AVG columns. In the raw database, if uncertainty was reported in the source publication, then a UNC column was included; there are no AVG columns. The legacy and raw databases include the id column described below, with assignments illustrated in Table S4.

Table S2: List of tables comprising the legacy database (ldb). Abbreviations were assigned in this work and used in subsequent tables to denote fire or fuel type.		
Table Name	Source (Akagi et al., 2011 supplementary tables)	Pollutant Categories
ldb_savanna (sv)	S1. Savanna	inorganic gas, methane, NMOC_g, PM
ldb_boreal_forest (bf)	S2. Boreal forest	inorganic gas, methane, NMOC_g, PM
ldb_tropical_forest (trf)	S3. Tropical forest	inorganic gas, methane, NMOC_g, PM
ldb_temperate_forest (tmf)	S4. Temperate forest	inorganic gas, methane, NMOC_g, PM
ldb_peat (p)	S5. Peat	inorganic gas, methane, NMOC_g, PM
ldb_chaparral (chp)	S6. Chaparral	inorganic gas, methane, NMOC_g, PM
ldb_open_cooking (ocook)	S7. Open cooking	inorganic gas, methane, NMOC_g, PM
ldb_cookstove (cs)	S8. Cookstove	inorganic gas, methane, NMOC_g, PM
ldb_dung_burning (db)	S9. Dung burning	inorganic gas, methane, NMOC_g, PM
ldb_charcoal_making (chrm)	S10. Charcoal making	inorganic gas, methane, NMOC_g, PM
ldb_charcoal_burning (chrb)	S11. Charcoal burning	inorganic gas, methane, NMOC_g, PM

ldb_pasture_maintenance (pm)	S12. Pasture maintenance	inorganic gas, methane, NMOC_g, PM
ldb_crop_residue (cr)	S13. Crop residue	inorganic gas, methane, NMOC_g, PM
ldb_garbage_burning (gb)	S14. Garbage burning	inorganic gas, methane, NMOC_g, PM

Table S3: Column headers and descriptions for tables in the legacy database (ldb).	
Column Name	Description
mm	Molar mass
formula	Chemical formula
compound	Compound name (as reported in the source publication)
pollutant_category	Categorizes the type of pollutant as either inorganic gas, methane, gaseous non-methane organic compound (NMOC_g), particulate non-methane organic compound (NMOC_p) or particulate matter (PM).
EF	All emission factor column names start with the string 'EF_akagi11_' and are then combined with the name from Akagi et al. (2011) that indicates fuel type and original data source; for example, 'EF_akagi11_boreal_organic_soil_bertschi03'.
UNC	UNC is the uncertainty reported in Akagi et al. (2011). The column name is analogous to the EF column name, where 'EF' is replaced with 'UNC'; for example, 'UNC_akagi11_boreal_organic_soil_bertschi03'. UNC is retained for individual studies and averaged across fuel/fire types as reported in Akagi et al. (2011).
AVG	Average is the weighted average EF reported in Akagi et al. (2011) for each fuel/fire type. The column name is analogous to the EF column name, where 'EF' is replaced 'AVG'; for example, 'AVG_akagi11_boreal_forest'.
id	Unique identifier assigned in the development of NEIVA

In order to create a merged dataset and link tables within NEIVA, each organic compound was assigned a unique identifier (id), including unidentified organic compounds designated by a formula. Unique ids were assigned through a hierarchical process, illustrated in Figure S1, in which preference was given to the use of the IUPAC international chemical identifier (InChI). InChI were assigned using Python packages (e.g., PubChemPy) and in-house web scrapers to access the online databases PubChem (Kim et al., 2021), NIST (NIST Chemistry WebBook, 2022), and ChemSpider (Pence and Williams, 2010) from which the InChI were obtained (Type 1 id). In some cases, EFs were reported for a group of identified compounds rather than individual compounds. For such cases, the unique ID was assigned as the exact mass and the chemical formula (Type 2 id). Finally, in some cases, EFs were reported for compounds with a known chemical formula but an unknown structure, and thus an InChI could not be assigned. For such compounds, the analytical method was used to generate the unique ID (Type 4 id). For compounds or groups of compounds detected using FTIR or PTR-TOF-MS, the unique ID was the exact or protonated mass (depending on what was reported in the source publication) and the chemical formula (e.g., 'butenes' was assigned the ID '56.0626_C4H8'). For compounds detected using GC×GC-TOF-MS, a combined string of the first- and second-dimension retention times was used to create the unique ID (e.g., a C₁₁H₁₄ isomer was assigned the ID '1769.44_1.729'). Table S4 illustrates each of the types of unique IDs assigned. Table S5 provides the list of tables that comprise the raw database.

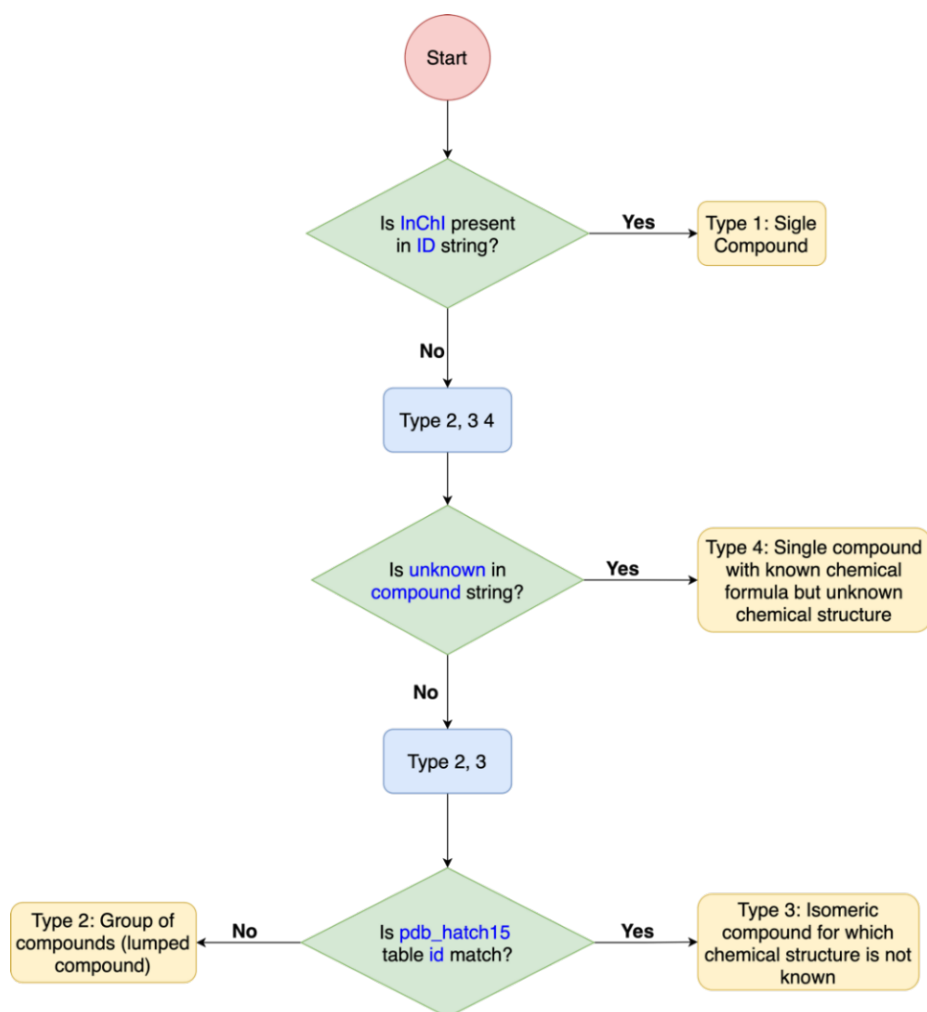


Figure S1: Flowchart illustrating hierarchical process for assigning unique identifiers (id) to organic compounds in NEIVA.

Table S4: Types of unique id with examples.			
Type	Example		
	Formula	Compound/ Compound Class	ID
Type 1: Single compound classified using InChI.	C ₁₀ H ₁₆	Camphene	InChI=1S/C ₁₀ H ₁₆ /c1-7-8-4-5-9(6-8)10(7,2)3/h8-9H,1,4-6H2,2-3H3
Type 2: Group of compounds ('lumped compound') classified using exact or protonated mass + formula.	C ₁₀ H ₁₆	Monoterpenes	136.12416000000002_C ₁₀ H ₁₆
Type 3: One or more isomeric compounds for which chemical structure is not known, classified using first and second retention time.	C ₁₀ H ₁₆	C ₁₀ H ₁₆ isomers	1069.66_1.091
Type 4: Single compound with known chemical formula but unknown chemical structure.	C ₄ H ₅ O	unknown	69.03346_C ₄ H ₆ O

Table S5: List of tables comprising the raw database (rdb). Abbreviations refer to one of the 14 fuel or fire types as defined in Table S2. For publications that include multiple fuel or fire types, no abbreviation for the fuel or fire type is given. Each table includes data from a single manuscript.

Fire Type	Table Name	Pollutant Categories	Source	DOI
Savanna	rdb_sv_desservettaz20	inorganic gas, methane, PM, NMOC_p	Table 4, Column: This Study	doi.org/10.1002/2016JD025925
Boreal forest	rdb_bf_hayden22	inorganic gas, NMOC_g, PM	Table A1	doi.org/10.5194/acp-22-12493-2022
Tropical forest	rdb_trf_hodgson18	inorganic gas, methane, PM	Table 3 Row- Rondonia This Study, Table 4 Row- Rondonia This Study, Tocantis This Study	doi.org/10.5194/acp-18-5619-2018
Temperate forest	rdb_tmf_gkatzelis23	inorganic gas, methane, NMOC_g, PM	Table 3	doi.org/10.5194/egusphere-2023-1439
	rdb_tmf_permar21	inorganic gas, methane, NMOC_g, PM	Table 2	doi.org/10.1029/2020JD033838
	rdb_tmf_liu17	inorganic gas, methane, NMOC_g, PM	Table 3	doi.org/10.1002/2016jD026315
	rdb_tmf_muller16	inorganic gas, NMOC_g	Table 2, Table 3	doi.org/10.5194/acp-16-3813-2016
Peat	rdb_p_watson19	inorganic gas, methane, NMOC_g, PM, NMOC_p	Table 2, Table 3, Table 4	doi.org/10.5194/acp-19-14173-2019
	rdb_p_jayarathne18	PM, NMOC_p	Table 2, 3.2 Emission of OC, EC and WSOC	doi.org/10.5194/acp-18-2585-2018
	rdb_p_roulston18	PM	Recommended data from 'Discussion and Conclusion'	doi.org/10.1029/2017JD027827

	rdb_p_smith17	inorganic gas, methane, NMOC_g	Table 3	doi.org/10.1002/2017GB005709
	rdb_p_stockwell16	inorganic gas, methane, NMOC_g, PM	Table S2	doi.org/10.5194/acp-16-11711-2016
Crop residue	rdb_cr_lasko18	PM	Table 2	doi.org/10.1016/j.envpol.2018.01.098
	rdb_cr_holder17	inorganic gas, methane, NMOC_g, PM	Table 3, Table S5, Table S2, Table S3 and Table 5	doi.org/10.1016/j.atmosenv.2017.06.043
	rdb_cr_liu16	inorganic gas, NMOC_g, PM	Table 3	doi.org/10.1002/2016JD025040
Garbage burning	rdb_gb_yokelson13	inorganic gas, methane, NMOC_g, PM	Table S1	doi.org/10.5194/acp-13-89-2013
Multiple fuel or fire types	rdb_travis23	inorganic gas, methane, NMOC_g, PM, NMOC_p	Table S1-EFs	doi.org/10.1029/2023JD039309
	rdb_fleming18	inorganic gas, methane, NMOC_g, PM	Table 1	doi.org/10.5194/acp-18-15169-2018
	rdb_goetz18	PM	Supplement section 3 and 4	doi.org/10.5194/acp-18-14653-2018
	rdb_jayarathne18	PM, NMOC_p	Table 2, 3.2 Emission of OC, EC and WSOC	doi.org/10.5194/acp-18-2585-2018
	rdb_koss18	inorganic gas, NMOC_g	S3 Emission Factors	doi.org/10.5194/acp-18-3299-2018
	rdb_selimovic18	inorganic gas, methane, NMOC_g, PM	Table S2	doi.org/10.5194/acp-18-2929-2018
	rdb_coffey17	inorganic gas, PM	Table S2	doi.org/10.1021/acs.est.7b02436

	rdb_hatch17	inorganic gas, methane, NMOC_g	Supplemental Table	doi.org/10.5194/acp-17-1471-2017
	rdb_pokhrel16	PM	Table S2	doi.org/10.5194/acp-16-9549-2016
	rdb_stockwell16	inorganic gas, methane, NMOC_g, PM	Table S8, Table S7, Table S9, Table 6	doi.org/10.5194/acp-16-11043-2016
	rdb_hatch15	inorganic gas, NMOC_g	Table S1	doi.org/10.5194/acp-15-1865-2015
	rdb_stockwell15	inorganic gas, methane, NMOC_g	Table S2	doi.org/10.5194/acp-15-845-2015
	rdb_jayarathne14	PM	Table 1	doi.org/10.1021/es502933j
	rdb_may14	PM	Table 3, Table 4	doi.org/10.1002/2014JD021848

^aOC = organic carbon, EC = elemental carbon, WSOC = water soluble organic carbon, TC = total carbon, eBC = equivalent black carbon, OBTF = open burn test facility

For laboratory studies in which individual fuels were burned, these fuels were mapped to one of the 14 fuel or fire types as summarized in Table S6. Domestic biomass burning categories were revised from Akagi et al. (2011) and are summarized in Table S7.

Table S6: Mapping of individual fuels from FLAME-4 and FIREX laboratory studies to fire type.	
Fire or Fuel Type	Individual Fuel
Savanna	savanna grass
Boreal forest	black spruce
Temperate forest	ponderosa pine, bear grass, ceanothus, Douglas fir, Engelmann spruce, Jeffrey pine, juniper pine, loblolly pine, lodgepole pine, longleaf pine, sagebrush, subalpine fir, wiregrass
Chaparral	chamise, manzanita, shrubland
Crop residue	Alfalfa, hay (organic), rice straw, wheat straw, winter wheat, corn, soybean, Kentucky bluegrass, millet (Ghana)

Table S7: Domestic biomass burning categories.				
Fire Type	Study	Cookstove/Fuel	Measurement Type	MCE
Open cooking	Coffey et al 17	Three stone with wood	Field	0.933
	Akagi_11(Christian et al 10)		Field	0.949
	Akagi_11(Roden et al 09)		Field	0.917
	Akagi_11(Johnson et al 08)		Field	0.949
	Akagi_11(Roden et al 06)		Field	0.896
	Akagi_11(Bertschi et al 03)		Field	0.910

	Akagi_11(Zhang et al 00)		Field	0.934
	Akagi_11(Brocard et al 96)		Field	0.930
	Stockwell et al	Three stone with twig	Lab	0.955
	Stockwell et al 16	Three stone with hardwood	Lab	0.955
	Stockwell et al 15	Three stone with hardwood	Lab	0.968
	Akagi_11(Smith et al 00)		Lab	0.937
Cookstove	Fleming et al 18	Chulha with brushwood	Field	0.937
	Fleming et al 18	Chulha with mixed fuel	Field	0.892
	Coffey et al 17	Gayapa with wood	Field	0.945
	Coffey et al 17	Philip with wood	Field	0.958
	Stockwell et al 16	Mudstove with wood	Field	0.933
	Stockwell et al 16	Biogas stove	Field	0.999
	Stockwell et al 16	Mudstove with wood	Field	0.914
	Akagi_11(Christian et al 10)	Patsari	Field	0.970
	Akagi_11(Johnson et al 08)	Patsari	Field	0.950
	Stockwell et al 16	Mudstove with wood	Lab	0.966
	Stockwell et al 16	Forced draft stove with wood	Lab	0.975
	Stockwell et al 16	Envirotek stove with hardwood	Lab	0.984
	Stockwell et al 16	Chimney stove with hardwood	Lab	0.983
	Stockwell et al 16 (biogas stove)	Biogas stove	Lab	0.954
	Stockwell et al 16	Biobriquette	Lab	0.985
	Stockwell et al 15	Envirofit rocket stove)	Lab	0.975
	Stockwell et al 15	Ezystove	Lab	0.968
Dung burning	Fleming et al 18	Chulha with pure dung	Field	0.865
	Fleming et al 18	Agithi with pure dung	Field	0.819
	Stockwell et al 16	Mudstove with dung ,hardwood	Field	0.912
	Stockwell et al 16	Open burning with dung	Field	0.876
	Stockwell et al 16	Mudstove with pure dung	Field	0.908
	Akagi_11(Christian et al 07)		Field	0.836
	Koss et al 18	Open burning with dung	Lab	0.899
	Selimovic et al 18	Open burning with dung	Lab	0.899
	Stockwell et al 16	Chimney stove with dung, hardwood	Lab	0.965
	Stockwell et al 16	Mudstove with dung, twig	Lab	0.980
	Stockwell et al 16	Three stone with dung	Lab	0.964

	Stockwell et al 16	Mudstove with dung	Lab	0.956
	Stockwell et al 16	Envirotek stove with dung, hardwood	Lab	0.971
	Stockwell et al 16	Chimney stove with dung, twig	Lab	0.957
	Stockwell et al 16	Mudstove with dung, hardwood	Lab	0.976
	Akagi_11(Keene et al 06)		Lab	0.844
Charcoal burning	Coffey et al 17	Philip with charcoal	Field	0.939
	Coffey et al 17	Coalpot with charcoal	Field	0.880
	Stockwell et al 16	Clamp kiln	Field	0.950
	Stockwell et al 16	Zigzag kiln	Field	0.994
	Stockwell et al 16	Zigzag kiln with stoke holes	Field	0.861
	Akagi_11(Bertschi et al 03)		Field	0.919
	Akagi_11(Kituyi et al 01)		Field	0.927
	Akagi_11(Brocard et al 98)		Field	0.866
	Stockwell et al 16	Forced draft stove	Lab	0.929
	Akagi_11(Smith et al 00)		Lab	0.848

S2. Primary Database

The data tables from the legacy and raw databases were reformatted to achieve a consistent structure that allowed combining the data into a single database, referred to as the NEIVA primary database (pdb). Data processing steps, described below, were performed on the legacy and raw databases prior to inclusion in the primary database. The resultant NEIVA primary database consists of 44 tables listed in Table S8.

Table S8: Tables comprising the primary database (pdb).		
Fire Type	Table Name	Source
savanna	pdb_akagi11_savanna	legacy DB (ldb)
boreal forest	pdb_akagi11_boreal_forest	legacy DB (ldb)
tropical forest	pdb_akagi11_tropical_forest	legacy DB (ldb)
temperate forest	pdb_akagi11_temperate_forest	legacy DB (ldb)
peat	pdb_akagi11_peat	legacy DB (ldb)
chaparral	pdb_akagi11_chaparral	legacy DB (ldb)
open cooking	pdb_akagi11_open_cooking	legacy DB (ldb)
cookstove	pdb_akagi11_cookstove	legacy DB (ldb)
dung burning	pdb_akagi11_dung_burning	legacy DB (ldb)
charcoal making	pdb_akagi11_charcoal_making	legacy DB (ldb)
charcoal burning	pdb_akagi11_charcoal_burning	legacy DB (ldb)
pasture maintenance	pdb_akagi11_pasture_maintenance	legacy DB (ldb)
crop residue	pdb_akagi11_crop_residue	legacy DB (ldb)
garbage burning	pdb_akagi11_garbage_burning	legacy DB (ldb)
savanna	pdb_sv_desservettaz20	raw DB (rdb)

boreal forest	pdb_bf_hayden22	raw DB (rdb)
tropical forest	pdb_trf_hodgson18	raw DB (rdb)
temperate forest	pdb_tmf_gkatzelis23	raw DB (rdb)
temperate forest	pdb_tmf_permar21	raw DB (rdb)
temperate forest	pdb_tmf_liu17	raw DB (rdb)
temperate forest	pdb_tmf_muller16	raw DB (rdb)
peat	pdb_p_watson19	raw DB (rdb)
peat	pdb_p_jayarathne18	raw DB (rdb)
peat	pdb_p_roulston18	raw DB (rdb)
peat	pdb_p_smith17	raw DB (rdb)
peat	pdb_p_stockwell16	raw DB (rdb)
crop residue	pdb_cr_lasko18	raw DB (rdb)
crop residue	pdb_cr_holder17	raw DB (rdb)
crop residue	pdb_cr_liu16	raw DB (rdb)
garbage burning	pdb_gb_yokelson13	raw DB (rdb)
multiple fire types	pdb_travis23	raw DB (rdb)
multiple fire types	pdb_fleming18	raw DB (rdb)
multiple fire types	pdb_goetz18	raw DB (rdb)
multiple fire types	pdb_jayarathne18	raw DB (rdb)
multiple fire types	pdb_koss18	raw DB (rdb)
multiple fire types	pdb_selimovic18	raw DB (rdb)
multiple fire types	pdb_coffey17	raw DB (rdb)
multiple fire types	pdb_hatch17	raw DB (rdb)
multiple fire types	pdb_pokhrel16	raw DB (rdb)
multiple fire types	pdb_stockwell16	raw DB (rdb)
multiple fire types	pdb_hatch15	raw DB (rdb)
multiple fire types	pdb_stockwell15	raw DB (rdb)
multiple fire types	pdb_jayarathne14	raw DB (rdb)
multiple fire types	pdb_may14	raw DB (rdb)

Data processing for legacy database:

- The legacy database includes EFs and uncertainty from Akagi et al. (2011). The uncertainty columns were not included in the primary database and were not used for calculating averages.
- The ‘akagi11_temperate_forest’ table included EFs from Radke et al. (1991) that were based on field measurements or estimates when field data were not available. The estimates, an 80/10/10 weighted average of pine-forest understory, coniferous canopy, and organic soil reported by Yokelson et al. (2013) were not included in the primary database.
- The ‘akagi11_temperate_forest’ and ‘akagi11_chaparral’ tables include EFs for unknown PIT masses. These unknowns were not retained in the primary database.
- The ‘akagi11_temperate_forest’ and ‘akagi11_chaparral’ tables included isomers as separate compounds; for example, ‘Other C₆H₁₀ (isomer_1)’ and ‘Other C₆H₁₀

(isomer_2)'. In the primary database isomers were combined, renamed (e.g., 'C6H10 isomers'), and the EFs summed over all isomers.

- The 'akagi11_peat' table in the legacy database includes peatland data from Yokelson et al. (1997) and Christian et al. (2003). Because peatland typically includes peat soil and the overlaying vegetation, only the peat data were retained in the primary database, thus omitting the peatland data and use of the "peatland" descriptor.
- Two conventions are common for reporting charcoal-making EFs: grams per kg of wood used and grams per kg of charcoal made, where a kg of wood typically yields about 280 g of charcoal (Bertschi et al., 2003). In Tables 2 and S9 of Akagi et al. (2011), the factor to convert g/kg-wood to g/kg-charcoal-made (~3.57) was applied to all the gases when needed, but inadvertently omitted for EC and OC. That oversight has been corrected in the pdb.

Data processing for raw database:

- For studies in which multiple fires of the same type were sampled, only the average was retained in the primary database. For example, the 'rdb_p_stockwell16' table has EF data for 35 peat fire samples and only the average EF of those 35 samples was retained in the pdb.
- In the 'rdb_stockwell15' table, zeros were added for peat nitrogen oxide (NO), nitrogen dioxide (NO₂), nitrous acid (HONO), and sulfur dioxide (SO₂) measurements that were below the detection limit (in 4-5 of 6 peat stack burns depending on the constituent) and the zeros were used to calculate the average EFs for those constituents in the pdb for peat from Stockwell et al. (2015).
- From the 'rdb_hatch17' table, only the WAS data were retained in the primary database since the other EF data were included in 'hatch15' or 'stockwell15'.
- The 'rdb_hatch15' table included isomers as separate compounds. If isomers belonged to the same functional group type, in the primary database they were combined, renamed, and the EF was summed over all isomers. If they belonged to different functional group types, they remained separate. For example, a 'C11H16 isomer' of type "aliphatics" and of type "aromatics-0 DBE" would be represented separately (2 rows) in the primary database: 'C11H16 isomers', one with VOC type 'aliphatics-other' and the other with VOC type 'aromatics- 0 DBE' (where DBE = double bond equivalent).
- The units of PM_{2.5} and OC from 'rdb_p_jayarathne18' were converted from mg/g to g/kg before inclusion in the primary database. Similarly, the weight percent of PM_{2.5} and OC in mg/g in 'jayarathne18_n' were converted to g/kg before being retained in the primary database.
- To avoid double counting, the EF data for elemental carbon (EC) was dropped from the 'pdb_p_jayarathne18' dataset because the same EF data for EC was reported in 'pdb_p_stockwell16'.
- EFs for gaseous NMOCs measured during FIREX and reported by Koss et al. (2018) were calculated assuming 50 % carbon for all samples. Prior to their inclusion in the primary database, the EFs were multiplied by a fuel-dependent correction factor to reflect the actual % carbon reported by Selimovic et al. (2018). The % C correction factors, summarized in Table S9, were derived by taking the ratio of the measured % C/assumed % C (50) averaged for each fuel type.

Table S9: Average EFs based on Koss et al. (2018) and % C correction factor applied to account for actual % C reported by Selimovic et al. (2018).	
Average EF	% C correction factor
EF_engelmann_spruce_koss18	0.95
EF_jeffrey_pine_koss18	0.98
EF_sagebrush_koss18	0.93
EF_manzanita_uncontaminated_koss18	0.96
EF_manzanita_contaminated_koss18	0.96
EF_bear_grass_koss18	0.92
EF_rice_straw_koss18	0.79
EF_peat_koss18	1.14
EF_dung_koss18	0.75
EF_excelsior_koss18	0.97
EF_akagill_crop_residue_yokelson11	0.90

S3. Integrated EF Dataset

The study-averaged EF data aggregated in the NEIVA primary database were merged into a single dataset, the NEIVA integrated EF dataset, which consists of a single EF table for all fuel and fire types with the columns as listed in Table S10. The multistep process for merging the EF data is described below.

Table S10: Columns of the NEIVA integrated dataset.	
Column name	Description
mm	Molar mass
formula	Chemical formula
compound	Chemical compound name
pollutant_category	Categorizes the type of pollutant as either inorganic gas, methane, gaseous non-methane organic compound (NMOC_g), particulate non-methane organic compound (NMOC_p), or particulate matter (PM).
EF	Columns imported from the legacy database are prefixed with 'EF_akagill_'. Columns imported from the raw database follow the general format 'EF_[Fuel Type]_[Study]'; e.g., 'EF_ponderosa_pine_hatch15'.
id	Unique identifier assigned in the development of NEIVA

The primary database tables were merged into a single table using a hierarchical process based on the unique ids assigned in this work and executed using a Python script. First, compounds or groups of compounds with matching 'id' and with unmatching 'id' were identified across the tables in the primary database. Second, for matched ids EF columns were appended to a merged table (indicating additional EFs), while for unmatched id rows were appended (indicating a new compound). This two-step process was applied iteratively to each of the pdb tables to create a single merged table.

One of the limitations in the above approach is that grouped compounds of the same compound class but with different names were identified as unique compound classes, leading to double counting in subsequent calculations. To resolve this, after the two-step process, additional logical conditions were applied. Specifically, all chemical formulas that had more than one id based on

exact/protonated mass or retention time were flagged. This resulted in 60 flagged groups of compounds. A subset of the flagged compounds is shown in Table S11. Rows with the same formula and equivalent ids were combined into a single row in the merged table (EF columns for each entry appended). The compound name with the longest string was chosen as the representative name and its id was adopted for that row in the merged table. Table S12 illustrates the results of these combinations for the subset of compounds in Table S11. The number of NMOC_g compounds (rows in the table) pre and post this additional refinement is listed in Table S13.

After the final merge, compounds were sorted. Gaseous compounds ('NMOC_g', 'inorganic gas', 'methane') were arranged in ascending order by molar mass. Particulate compounds were arranged with total PM followed by PM constituents (e.g., OC, EC, BC, NMOC_p and optical property). In the final integrated EF dataset, the compounds were ordered with methane followed by inorganic gas, then NMOC_g, and finally PM (including PM constituents). The integrated EF dataset has a total of 1311 rows (i.e., compounds or constituents) and 263 columns (i.e., individual EFs). In other words, EFs are available for a total of 1311 compounds or constituents with up to 263 individual or study-averaged measurements across the 14 major fuel and fire types.

Table S11: A subset of the integrated EF dataset that represents groups of compounds with equivalent ids but different names.			
Formula	Compound Name	ID	Source Table in PDB
C10H16	Monoterpenes	136.12416000000002_C10H16	pdb_tmf_permar21, pdb_koss18, pdb_tmf_liu17, pdb_akagi11_temperate_forest pdb_akagi11_chaparral
C10H16	Terpenes	136.124677_C10H16	pdb_tmf_gkatzelis23, pdb_travis23, pdb_bf_hayden22, pdb_stockwell15
C7H8O	Cresols(Methoxyphenols)	108.056991_C7H8O	pdb_stockwell15
C7H8O	2-Methylphenol (=o-cresol) + anisol	108.05716_C7H8O	pdb_tmf_gkatzelis23, pdb_travis23, pdb_bf_hayden22, pdb_tmf_permar21, pdb_koss18,
C8H10	C8 Aromatics	106.077687_C8H10	pdb_bf_hayden22, pdb_akagi11_temperate_forest
C8H10	Ethyl benzene + m-xylene + p-xylene + o-xylene	106.07816_C8H10	pdb_tmf_gkatzelis23, pdb_koss18, pdb_stockwell15, pdb_akagi11_savanna, pdb_akagi11_tropical_forest, pdb_akagi11_pasture_maintenance
C8H10	C8 Aromatics	106.07825_C8H10	pdb_tmf_permar21

C8H10	m+p-xylene	106.078250319_C8H10	<p> pdb_travis23, pdb_tmf_permar21, pdb_fleming18, pdb_holder17, pdb_hatch17, pdb_tmf_liu17, pdb_stockwell16, pdb_hatch15, pdb_akagi11_boreal_forest pdb_akagi11_temperate_forest pdb_akagi11_chaparral </p>
-------	------------	---------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 12: The subset of grouped compounds from Table S11 after non-unique groups of compounds were combined as described in text.

Formula	Compound Name	ID	Source Table in PDB
C10H16	Monoterpenes	136.12416000000002_C10H16	<p> pdb_tmf_gkatzelis23, pdb_travis23, pdb_bf_hayden22, pdb_tmf_permar21, pdb_koss18, pdb_tmf_liu17, pdb_stockwell15 pdb_akagi11_temperate_forest pdb_akagi11_chaparral </p>
C7H8O	2-Methylphenol (=o-cresol) + anisol	108.05716_C7H8O	<p> pdb_tmf_gkatzelis23, pdb_travis23, pdb_bf_hayden22, pdb_tmf_permar21, pdb_koss18, pdb_stockwell15 </p>
C8H10	Ethyl benzene + m-xylene + p-xylene + o-xylene	106.07816_C8H10	<p> pdb_tmf_gkatzelis23 pdb_travis23, pdb_bf_hayden22, pdb_tmf_permar21, pdb_fleming18, pdb_koss18, pdb_hatch17, pdb_holder17, pdb_stockwell16, pdb_stockwell15 pdb_hatch15, pdb_akagi11_savanna, pdb_akagi11_boreal_forest, pdb_akagi11_tropical_forest, pdb_akagi11_temperate_forest, pdb_akagi11_chaparral, pdb_akagi11_pasture_maintenance </p>

Table S13: The size of the integrated EF dataset through the integration steps.		
Dataset	Size (columns x rows)	Number of Unique Formulas
Initial integrated NMOC_g table, combination of all pdb tables using id column.	255 x 1189	591
Refined integrated NMOC_g table, after combination of non-unique groups of compounds.	255 x 1115	591
Final integrated EF table that includes inorganic gas, methane, NMOC_g, NMOC_p, and PM EFs for all fire and fuel types.	255 x 1296	708

S4. Processed EF Dataset

Two significant data processing steps were performed using the integrated EF dataset prior to calculating the averages for the recommended EF dataset. Laboratory-based EFs were adjusted to account for known differences in laboratory and field combustion conditions and isomer distributions were assigned to grouped compounds where applicable. Each of these steps are described in detail below and the resultant EFs are stored in the processed EF dataset, a single table for all EFs across all 14 fuel and fire types.

Laboratory-Based EF Adjustment

The laboratory-based EF data for all fuels, with the exception of peat, were adjusted to account for known differences in modified combustion efficiency (MCE) between laboratory and field studies. Specifically, MCE in laboratory burns is typically higher than in field measurements. Previous studies have described and applied methods for adjusting laboratory-derived EFs (e.g., Christian et al., 2003, Yokelson et al., 2008, Selimovic et al., 2018). Most commonly, laboratory EFs for individual compounds or classes of compounds are plotted as a function of MCE and the data are fit using linear regression; the slope and intercept of the linear fit allows calculation of a field-adjusted EF based on the field-derived MCE. This method requires having enough data points for each compound or class of compounds to obtain a robust linear regression.

Here, a modified approach was applied using emission ratios (ERs) to CO (smoldering-dominant compounds) and CO₂ (flaming-dominant compounds). For each study, the average EF was calculated for all fuels mapped to a single fuel or fire type. For example, the pdb_koss18 table includes EFs for individual fuels such as ponderosa pine, Engelmann spruce, chamise, and peat. For fuels that were mapped to temperate forest (e.g., ponderosa pine, Engelmann spruce, see Table S6) the calculated average was stored in the processed EF dataset with the column name 'EF_temperate_forest_koss18'. Averaging across fuels mapped to a specific fire type provides a better representation of mixed fuels encountered in the field. For each gaseous compound (*i*) in each study, the average ER was then calculated relative to the study-average CO for smoldering-dominant compounds and CO₂ for flaming-dominant compounds. The laboratory-based study-averaged ER was then multiplied by the field-averaged EF_CO to calculate a laboratory-adjusted EF as follows for smoldering-dominant compounds:

$$EF_{i_lab_adj} = (EF_{i_lab_ave}/EF_{CO_lab_ave}) * EF_{CO_field_ave}$$

and by the field-averaged EF_CO₂ for flaming-dominant compounds:

$$EF_{i_lab_adj} = (EF_{i_lab_ave}/EF_{CO2_lab_ave}) * EF_{CO2_field_ave}$$

Table S14 summarizes the data used in the laboratory-based EF adjustment and the MCE values pre- and post-EF adjustment (see also Figure 1 and Figure S2).

Table S14: Reference publications used in the emission factor adjustment.				
Fire Type	Study	Measurement Type	MCE Pre-adjustment	MCE Post-adjustment
Savanna	Desservettaz et al. 20	Field	0.899	
	Akagi_11(Yokelson et al. 11)	Field	0.930	
	Akagi_11(Yokelson et al. 03)	Field	0.938	
	Akagi_11(Sinha et al. 03)	Field	0.941	
	Akagi_11(Ferek et al. 98)	Field	0.962	
	Travis et al. 23	Field	0.895	0.934
	Stockwell et al. 15	Lab	0.958	0.934
	Akagi_11(Christian et al. 03)	Lab	0.953	0.934
Boreal forest	Hayden et al. 22	Field	0.891	
	Akagi_11(Simpson et al. 11)	Field	0.901	
	Akagi_11(Goode et al. 00)	Field	0.921	
	Akagi_11(Nance et al. 93)	Field	0.928	
	Akagi_11(Radke et al. 91)	Field	0.911	
	Hatch et al. 17	Lab	0.937	0.911
	Hatch et al. 15	Lab	0.937	0.911
	Stockwell et al. 15	Lab	0.959	0.911
	Akagi_11(Burling et al. 10)	Lab	0.827	0.911
	Akagi_11(Bertschi et al. 03)	Lab	0.874	0.911
	Akagi_11(Yokelson et al. 97)	Lab	0.821	0.911
Temperate forest	Gkatzelis et al. 23	Field	0.899	
	Permar et al. 21	Field	0.901	
	Liu et al. 17	Field	0.912	
	Muller et al. 16	Field		
	Akagi_11(Yokelson et al. 13) organic soil	Field	0.850	
	Akagi_11(Yokelson et al. 13) coniferous canopy	Field	0.926	
	Akagi_11(Akagi et al. 13)	Field	0.931	
	Akagi_11(Yokelson et al. 13)	Field	0.936	
	Akagi_11(Yokelson et al. 11)	Field	0.908	
	Akagi_11(Radke et al. 91) debris	Field	0.927	

	Akagi_11(Radke et al. 91) temperate wildfire	Field	0.914	
	Travis et al. 23	Field	0.914	0.912
	Selimovic et al. 18	Lab	0.92	0.912
	Koss et al. 18	Lab	0.926	0.912
	Hatch et al. 17	Lab	0.927	0.912
	Hatch et al. 15	Lab	0.927	0.912
	Stockwell et al. 15	Lab	0.948	0.912
Chaparral	Akagi_11(Yokelson et al. 13)	Field	0.935	
	Akagi_11(Radke et al. 91)	Field	0.946	
	Travis et al. 23	Lab	0.939	0.941
	Koss et al. 18	Lab	0.958	0.941
	Selimovic et al. 18	Lab	0.958	0.941
	Stockwell et al. 15	Lab	0.929	0.941
Crop residue	Holder et al. 17	Field	0.951	
	Liu et al. 16	Field	0.93	
	Stockwell et al. 16	Field	0.952	
	Akagi_11(Yokelson et al. 11)	Field	0.925	
	Travis et al. 23	Field	0.929	0.946
	Koss et al. 18	Lab	0.953	0.946
	Selimovic et al. 18	Lab	0.953	0.946
	Hatch et al. 17	Lab	0.942	0.946
	Hatch et al. 15	Lab	0.942	0.946
	Stockwell et al. 15	Lab	0.941	0.946
	Akagi_11(Christian et al. 03)	Lab	0.811	0.946
Garbage burning	Stockwell et al. 16	Field	0.923	
	Akagi_11(Yokelson et al. 11)	Field	0.974	
	Akagi_11(Christian et al. 10)	Field	0.951	
	Stockwell et al. 15	Lab	0.973	0.948
	Yokelson et al. 13	Lab	0.967	0.948

The results of the laboratory-based EF adjustment are further explored and evaluated in Figures S2-S5 below, in which the magnitude of the EF adjustment for flaming-dominant compounds is shown (Figure S2), and comparisons are made between calculated averages using laboratory data only, field data only, and the average of laboratory-corrected and field data, with published EFs from recent field studies.

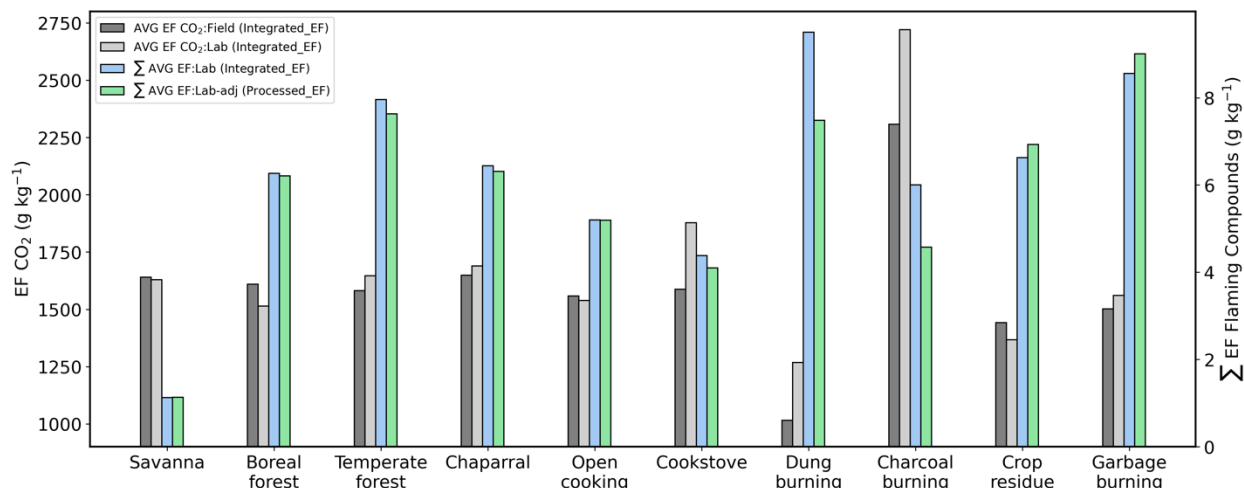


Figure S2: Averaged EF values for CO₂ (field, dark grey; lab, light grey) and the sum of flaming dominant compounds (excluding CO₂) pre-(blue) and post-(green) adjustment to account for differences in combustion conditions between laboratory and field.

Figure S2 illustrates the magnitude of the adjustment to laboratory-based EFs for flaming dominant compounds (NO, NO₂, NO_x as NO, N₂O, HONO, SO₂, HCl, gaseous Hg). For each fuel or fire type, the average field-based EF for CO₂ is shown in dark grey and the laboratory-based EF for CO₂ in light grey. The laboratory-based CO₂ values are higher for temperate forest, chaparral, domestic bb (cookstove, dung, charcoal), and garbage burning. The sum of the adjusted EFs for the flaming-dominant compounds decreases for those fuel and fire types, with the exception of garbage burning, to account for the higher EF values measured under more flaming conditions in the laboratory studies. For two fire types, boreal forest and garbage burning the sum of the adjusted EFs does not increase and decrease (respectively) as expected. This is likely because the natural variability (driven by fuel and fire characteristics) is larger than the small difference between the average field and laboratory EF_{CO₂}.

The adjusted laboratory-based EFs replace the unadjusted laboratory-based EFs in the processed EF dataset and are used in the calculation of the recommended EFs. To more closely evaluate this adjustment on an individual compound level, Figures S3-S5 show the distribution of field and adjusted laboratory EFs (box and whiskers) for the 25 compounds with the highest number of observations (“n”) in the NEIVA integrated EF database for the temperate forest fire type (Fig. S3), the corresponding figure for the crop residue fire type (Fig. S4), and the 25 most abundant NMOC_g in the crop residue fire type not shown in Fig. S4 (Fig. S5).

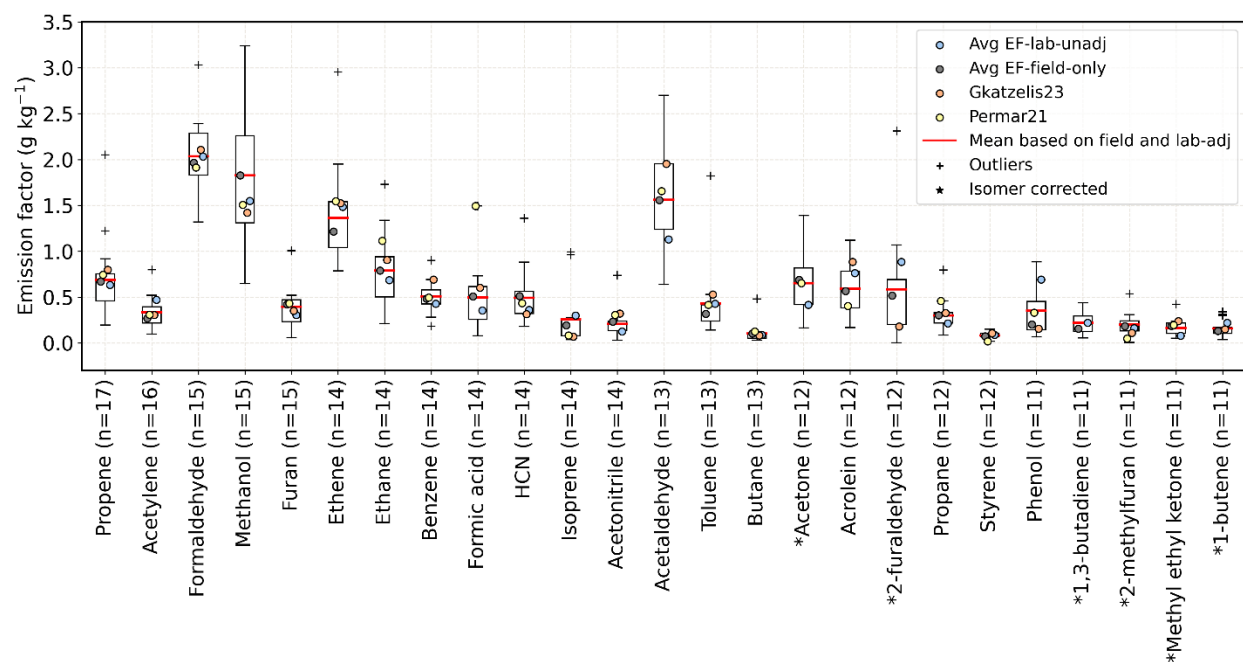


Figure S3: The NMOG_g EFs with the highest n (excluding compounds shown in Fig. 3) for temperate forest. The box and whiskers represent the values in the processed EF dataset and thus include the adjusted laboratory EFs. The red line indicates the mean value and is equivalent to the recommended EF. The stars represent outliers. Compounds marked with an asterisk have had an additional correction, application of isomeric distributions described below. The number of observations is listed in parenthesis (“n”).

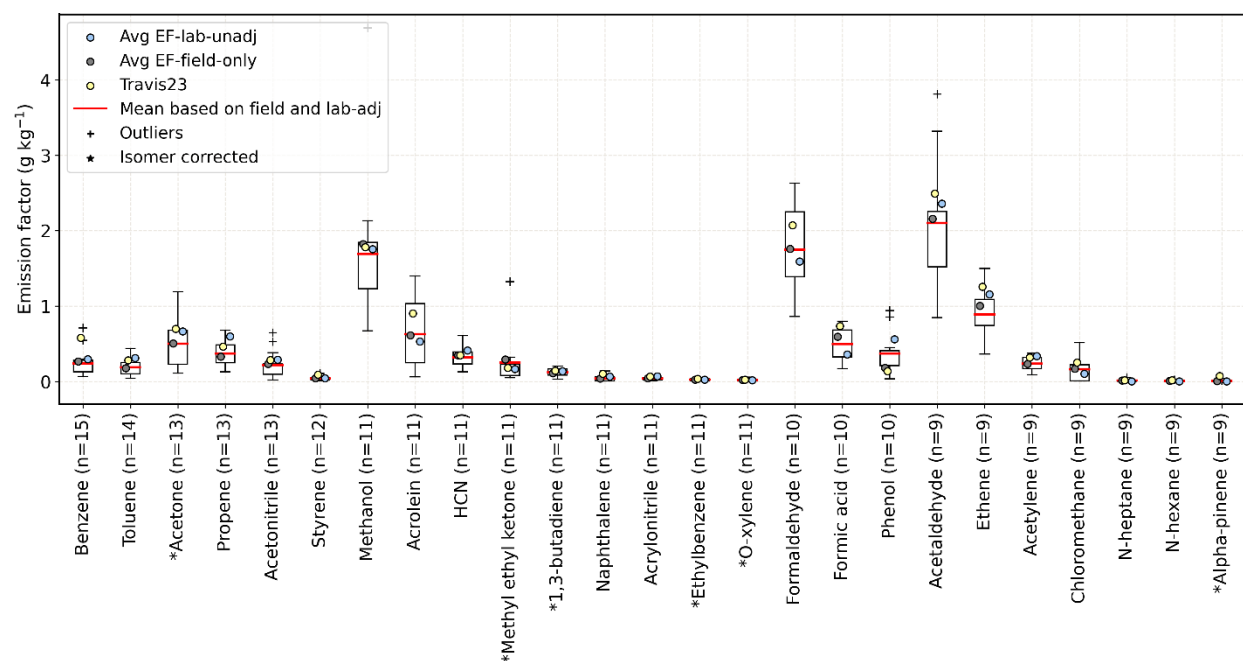


Figure S4: The NMOG_g EFs with the highest n (excluding compounds shown in Fig. S5) for crop residue. The box and whiskers represent the values in the processed EF dataset and thus include the adjusted laboratory EFs. The red line indicates the mean value and is equivalent to the recommended EF. The stars represent outliers. Compounds marked with an asterisk have had an additional correction, application of isomeric distributions described below. The number of observations is listed in parenthesis (“n”).

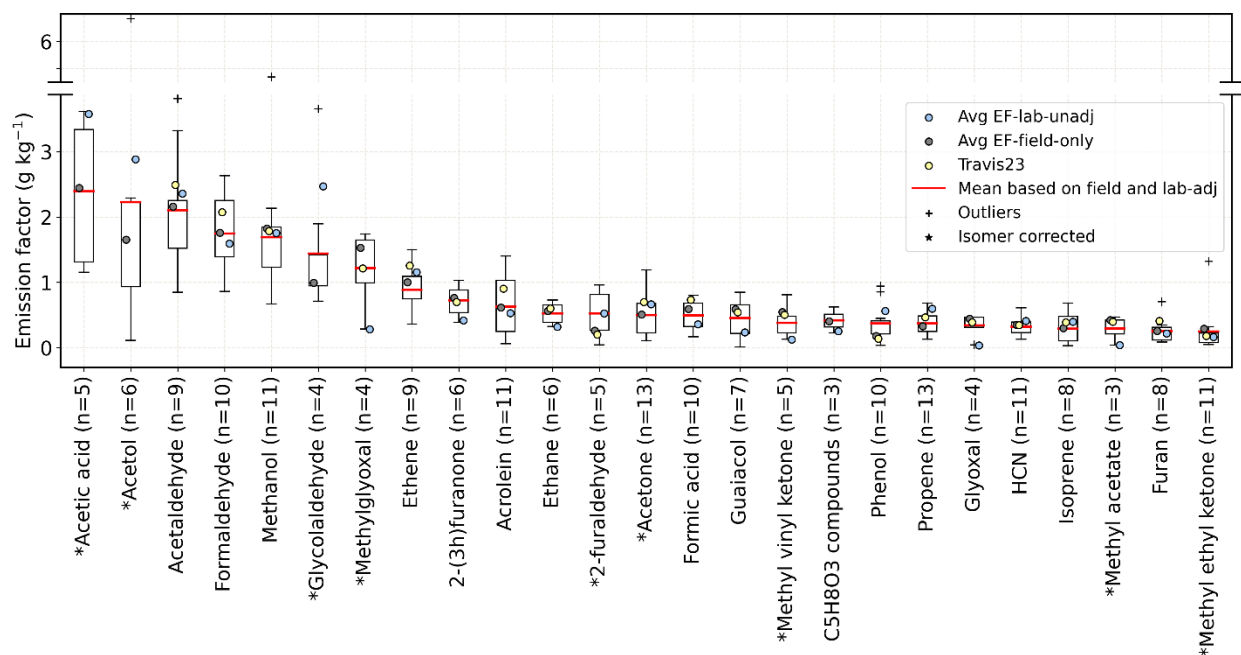


Figure S5: The 25 most abundant NMOC_g EFs for crop residue. The box and whiskers represent the values in the processed EF dataset and thus include the adjusted laboratory EFs. The red line indicates the mean value and is equivalent to the recommended EF. The stars represent outliers. Compounds marked with an asterisk have had an additional correction, application of isomeric distributions described below. The number of observations is listed in parenthesis (“n”).

Assignment of Fractional Contributions to Grouped Non-Methane Organic Compounds

For the NMOC_g in the integrated EF dataset, which represent the 14 fuel and fire types, speciated EF data were used to assign fractional contributions to compounds that could not be differentiated using the published method of detection. In general, if the sum of individual reported EFs was within $2.5 \times$ the EF for the grouped compounds (i.e., $\text{sum individual EFs} = 0.4\text{--}2.5 \times \text{grouped EF}$) in a given fuel or fire type, then the speciated EFs were used to assign fractional contributions to the grouped compounds. The only criteria used were the fuel or fire type and the EF range, thus the assignments were not restricted to e.g., samples collected from the same laboratory experiment or same fire plume. After applying the fractional contributions, the EF of the grouped compounds was removed from the dataset. If the sum of the individual reported compounds was outside $2.5 \times$ the EF for the grouped compound (i.e., $\text{sum individual EFs} = <0.4$ or $>2.5 \times \text{grouped EFs}$), the grouped compounds were retained as a single compound in the recommended EF database. This process minimizes double counting of NMOCs_g and allows better representation of the chemistry of individual compounds. To illustrate, in the integrated EF dataset the following results appear for “C₄H₆O” in a given fuel type. The first AVG column represents the average of all EFs included in the integrated EF dataset for that compound/group of compounds for that fuel or fire type. Note that this average is not weighted by the number of studies or measurements.

Formula	Compound	AVG EF
C ₄ H ₆ O	Methyl vinyl ketone + Methacrolein + Crotonaldehyde	0.2297
C ₄ H ₆ O	2-Butenal	0.0794
C ₄ H ₆ O	Methyl vinyl ketone	0.1348
C ₄ H ₆ O	Methacrolein	0.0562

The first row, 'Methyl vinyl ketone + Methacrolein + Crotonaldehyde', represents a group of compounds that could not be differentiated in the study/studies from which the average EF was calculated. For the same fuel type, EFs for 2-butenal (syn. crotonaldehyde), methyl vinyl ketone, and methacrolein were reported individually in another study/studies. Since the sum of the EFs for 2-butenal, methyl vinyl ketone, and methacrolein is within 2.5× the EF of 'Methyl vinyl ketone + Methacrolein + Crotonaldehyde', the relative distribution of the individual compounds is used to calculate a fractional contribution of each compound to the single EF, and the AVG is recalculated accordingly. In the resultant dataset, Methyl vinyl ketone + Methacrolein + Crotonaldehyde is removed.

Formula	Compound	AVG EF
C4H6O	Methyl vinyl ketone + Methacrolein + Crotonaldehyde	
C4H6O	2-Butenal	0.0734
C4H6O	Methyl vinyl ketone	0.1281
C4H6O	Methacrolein	0.0533

The criteria for this process are as follows: within a unique chemical formula, there must be both a 'lumped compound' (Category 2 id) and 'speciated compounds' (Category 1 id). In some cases, the lumped compound explicitly lists the speciated compounds, indicated by a '+' symbol in the name (e.g., '1-butyne + 2-butyne,' 'Acetic acid + glycolaldehyde,' 'Acetone + propanal'). The Python script identifies these by searching for the '+' symbol, splitting the compound name accordingly, and then searching for these individual ids in the merged dataset. If found, they are grouped together. Consequently, two distinct subsets of the dataset are used for the calculation of fractional contribution: one contains lumped compounds that explicitly list the individual compounds, while the other includes lumped compounds that do not specify the individual compounds. Table S15 lists the size of these subsets and the number of unique formulae in each, resulting in a total of 135 unique formula for which fractional distributions were assigned.

Table S15: Descriptions of NMOC_g datasets selected for fractional contribution assignment.		
Dataset	Size (columns x rows)	Number of Unique Formulae
Lumped NMOC_g compounds with speciated compounds	255 x 57	15
Lumped NMOC_g compounds without speciated compounds	255 x 540	120

S5. Recommended EF Dataset

The final product in the NEIVA database is a dataset of recommended EFs based on the study-averaged EFs summarized in the integrated EF dataset. The recommended EF dataset includes a single average EF for each compound or constituent in each of the 14 fuel and fire types, and an uncertainty estimate based on one standard deviation of the averaged EF values. Prior to averaging, NO_x EFs in the integrated EF dataset were converted to NO equivalent EFs as follows, when NO and NO₂ EFs were also available:

$$\text{EF NO}_x \text{ as NO} = \text{EF NO} + \text{EF NO}_2 * (\text{M}_{\text{NO}}/\text{M}_{\text{NO}_2}) [\text{M}_{\text{NO}}/\text{M}_{\text{NO}_2} = 30/46]$$

The columns in the recommended EF dataset are listed in Table S16.

Table S16: Column headers and rows in the recommended EF dataset.	
Column name	Description
mm	Molecular weight
formula	Chemical formula
compound	Chemical compound name
AVG	Calculated average of all emission factor columns within the specific fire type
N	Number of studies included in the average calculation
STD	Standard deviation
study	The first author's name and year of publication is combined in one string. Multiple studies are separated by ';'. For instance (stockwell15; hatch15)
id	Unique identity

S6. Chemical Mechanism Surrogate and Property Dataset

For the gaseous NMOCs included in the integrated, processed, and recommended EF datasets, model surrogate species were assigned for SAPRC-07/07T/18/22 (Carter, 2010, 2020, 2023a), MOZART-T1(Emmons et al., 2020), and GEOS-Chem (Bey et al., 2001; Carter et al., 2022) chemical mechanisms. The number of model surrogates used to represent the ~1000 individual gaseous NMOCs in NEIVA are summarized in Table S17 for each chemical mechanism.

Table S17: Number and list of model surrogates used to represent the individual gaseous NMOCs.		
Chemical Mechanism	Number of Surrogates	List of Model Surrogates
SAPRC-07, SAPRC-07T* *in S-07T the following species are explicitly represented: 1,3 butadiene (13BDE), acrolein (ACRO), alpha-pinene (APIN), 1,2,4-trimethylbenzene (B124), ethanol (ETOH), m-/o-/p-xylene (MXYL, OXYL, PXYL), propene (PRPE), and toluene (TOLU)	37	ACYE, NROG, ETHE, HCHO, ALK1, MEOH, OLE1, PRD2, ALK3, CCHO, ALK2, ALK5, FACD, OLE2, MACR, GLY, RCHO, ACET, AACD, ALK4, ARO2, ISOP, PACD, IPRD, MVK, MGLY, MEK, RNO3, BENZ, ARO1, BACL, CRES, BALD, TERP, ROOH, AFG3
SAPRC-18, SAPRC-22* *in S-22, there are four additional species: higher MW alkanes (ALK6) and amines bounded to a tertiary carbon (TAMNS) are added, catechol (CATL) replaces SVPHE, and acetic acid (OACID) replaces (AACID)	79	ACETL, NROG, ETHEN, HCHO, ETHAN, MEOH, OLE1, ACYLS, KET2, PROPE, AMINS, OTH3, MECHO, PROP, ETOH, HCOOH, OTH4, BUT13, ACRO, OLE2, OLE3, GLY, OLEP, ETCHO, ACET, NC4, ALK3, OTH1, AACID, GLCHO, ROOH, OLEC, FURNS, OLED, ISOP, RCOOH, OLEA1, MVK, MACR, ALK4, OLE4, LVKS, MGLY, RCHO, MEK, R1NO3, BENZ, ARO1, OLEA2, ALK5, BACL, OTH2, TOLU, PHEN, MALAH, STYRS, BALD, C2BEN, OXYL, MXYL, PXYL, CRES, XYNL, SVPHE, TERP, NAPS, ARO2, BZ124, BZ135, BZ123, AFG3, BENX, APINE, BPINE, DLIMO, INHIB, RTCHO, SESQ
MOZART-T1	37	C2H2, NROG, C2H4, C2H6, CH2O, CH3OH, BIGENE, C3H6,

		CH3CHO, C3H8, C2H5OH, HCOOH, GLYOXAL, BIGALK, CH3COCH3, CH3COOH, GLYALD, TOLUENE, ISOP, MVK, MACR, CH3COCHO, MEK, HYAC, BENZENE, XYLENES, MBO, PHENOL, ALKNIT, BZALD, CRESOL, MYRC, APIN, BPIN, LIMON
GEOS-Chem	36	ACET, ACTA, ALD2, BENZ, C2H4, C2H6, C3H8, CCl4, CH2Br2, CH3Br, CH3Cl, CH3I, CHBr3, CHCl3, DMS, EOH, ETNO3, GLYC, GLYX, HAC, HCOOH, ISOP, LIMO, MACR, MEK, MGLY, MOH, MTPA, MTPO, MVK, NAP, OCS, PROPNN, PYAC, TOLU, XYLE

Compounds were first assigned to the SAPRC and MOZART-T1 mechanisms (see Table S18) using the SAPRC Mechanism Generation (MechGen) System web interface (Carter, 2019, Carter, 2023b) and the SAPRC model species assignment database ‘SpecDB’(Carter, 2023b). The SAPRC and MOZART-T1 assignments were then used to determine the GEOS-Chem assignments (Table S19), with additional reference to reference to Hutzell et al. (2012), Li et al. (2014), and Carter et al. (2022).

Table S18: Mapping MOZART-T1 species to SAPRC-18 model species.	
SAPRC 18 Model Species	MOZART-T1 Model Species
HCHO	CH2O
KET2, LVKS, OLEA2, AFG1, MEK	MEK
OTH3, RCHO, ALK5	BIGALK
AMINS	C3H6
ACYLS, OTH4, OLEP, OLE1, OLE2, OLEC, OLED, BUT13, STYRS, OLE4	BIGENE
NROG, OTH2	NROG
ROOH, FURNS, ARO1, ARO2, NAPS	XYLENES
RCOOH, OLEA1	TOLUENE
RINO3	ALKNIT
PHEN, SVPHE	PHENOL
CRES, XYNL	CRESOL
DLIMO	LIMON
MVK	MVK

Table S19: Mapping GEOS-Chem model species to SAPRC-07 model species.	
SAPRC-07 Model Species	GEOS-Chem Model Species
ALK1	C2H6
ALK2	C3H8
ALK3, ALK4, ALK5, ACYE, PACD	ALK4
BALD	BALD
CRES	CSL
Among the CRES compounds, those identified as PHEN in S18B	PHEN
OLE1, OLE2	PRPE
RCHO	RCHO

ARO1	TOLU
ARO2	XYLE
TERP (excluding sabinene, alpha-pinene, beta-pinene, 3-carene)	MTPO
MVK, 0.5*IPRD	MVK
MACR, 0.5*IPRD	MACR
Among the ARO2 compounds, those identified as FURNS in S18B	FURA
Among the ARO2 compounds, those identified as NAPS in S18B	NAPS
BACL	MGLY
MEK	MEK
PRD2	MEK
ROOH	MP
TERP (excluding alpha pinene, beta pinene, sabinene, carene)	MTPO
RNO3	R4N2

Two SAPRC-07 model surrogates, AFG3 and NROG, were not assigned to GEOS-Chem species. The list of compounds that are mapped to those model surrogates is shown in Table S20. The summed EFs for compounds assigned to GEOS-Chem surrogates and for compounds that were not assigned to GEOS-Chem surrogates is shown in Table S21.

Table S20: The compounds mapped to SAPRC-07 species AFG3 and NROG, and thus not assigned in the GEOSChem EFs.			
mm	formula	compound	S07
27	HCN	Hydrogen cyanide	NROG
41	C2H3N	Acetonitrile	NROG
61	CH3NO2	Nitromethane	NROG
102	C4H6O3	Acetic anhydride	NROG
105	C3H7NO3	Isopropyl nitrate	NROG
117	C8H7N	Benzeneacetonitrile	NROG
119	C4H9NO3	2-Butyl nitrate	NROG
123	C6H5NO2	Nitrobenzene	NROG
124	C6H4O3	Hydroxybenzoquinone	AFG3
130	CF2Cl2	Dichlorodifluoromethane	NROG
131	C9H9N	Methyl benzeneacetonitrile	NROG
144	C2H3Cl3	1,1,1-Trichloroethane	NROG
147	CFC13	Trichlorofluoromethane	NROG
174	CHBrCl2	Bromodichloromethane	NROG
195	C2F3Cl3	Trichlorotrifluoroethane	NROG

Table S21: Summed EFs for compounds assigned and not assigned to GEOS-Chem model surrogates.		
Fire type	$\Sigma \text{EF}_{\text{NMOC}_g, \text{assigned}} \text{ (g/kg)}$	$\Sigma \text{EF}_{\text{NMOC}_g, \text{unassigned}} \text{ (g/kg)}$
Tropical forest	24.25	1.08
Temperate forest	40.39	2.10
Boreal forest	38.33	2.16
Savanna	34.43	2.85
Crop residue	36.71	1.35
Peat	65.75	7.9

In addition to the model surrogates, relevant property data for each of the NMOC_g in the integrated and recommended EF datasets was compiled in the property dataset. Most of the property data came from the Estimation Programs Interface (EPI) Suite developed by the US EPA (2023). In addition, a Python-based web scraper was used to collect data from the following data sources: UManSysProp (Topping et al., 2016), PubChem (Kim et al., 2021, 2023), ChemSpider (Pence and Williams, 2010), and NIST Chemistry Webbook (NIST Chemistry WebBook, 2022). When multiple values were available for a given property, experimental values were prioritized over the approximated values using the estimation methods. Data sources are listed in the reference column of the property dataset.

Table S22: The description and units of property variables.		
Column name	Description	Unit
mm	Molecular mass	g/mole
formula	Molecular formula	
compound	Compound name obtained from the source publication	
SMILES	Simplified Molecular Input Entry System	
S07	SAPRC-07 chemical mechanism model species	
S07T	SAPRC-07T (toxics version) chemical mechanism model species	
S18B	SAPRC-18 chemical mechanism model species	
S22	SAPRC-22 chemical mechanism model species	
MOZT1	MOZART-T1 chemical mechanism model species	
GEOSChem	GEOS-Chem chemical mechanism model species	
kOH	OH rate constant	cm ³ /molecule s
kOH_ref	Reference for OH rate constant data	
ko3_exp	O ₃ rate constant experimental values	cm ³ /molecule s
kno3_exp	NO ₃ rate constant experimental values	cm ³ /molecule s
vp_nannoolal	Vapor pressure estimated based on Nannoolal et al. (2008)	mm Hg
vp_EPISuite	Vapor pressure estimated based on EPI SUITE	mm Hg
vp_ref	Reference for vapor pressure data obtained from EPI SUITE	
cstar	Saturation vapor concentration (log 10)	μg/m ³
hc_EPISuite	Henry's law constant estimated based on EPI SUITE	atm-m ³ /mole
hc_ref	Reference for Henry's law constant obtained from EPI SUITE	
OCratio	O to C ratio	
Oxidation_state	Oxidation state	
id	Unique Identifier	

S7. Additional Results Figures

Comparisons between NEIVA and Andreae (2019)

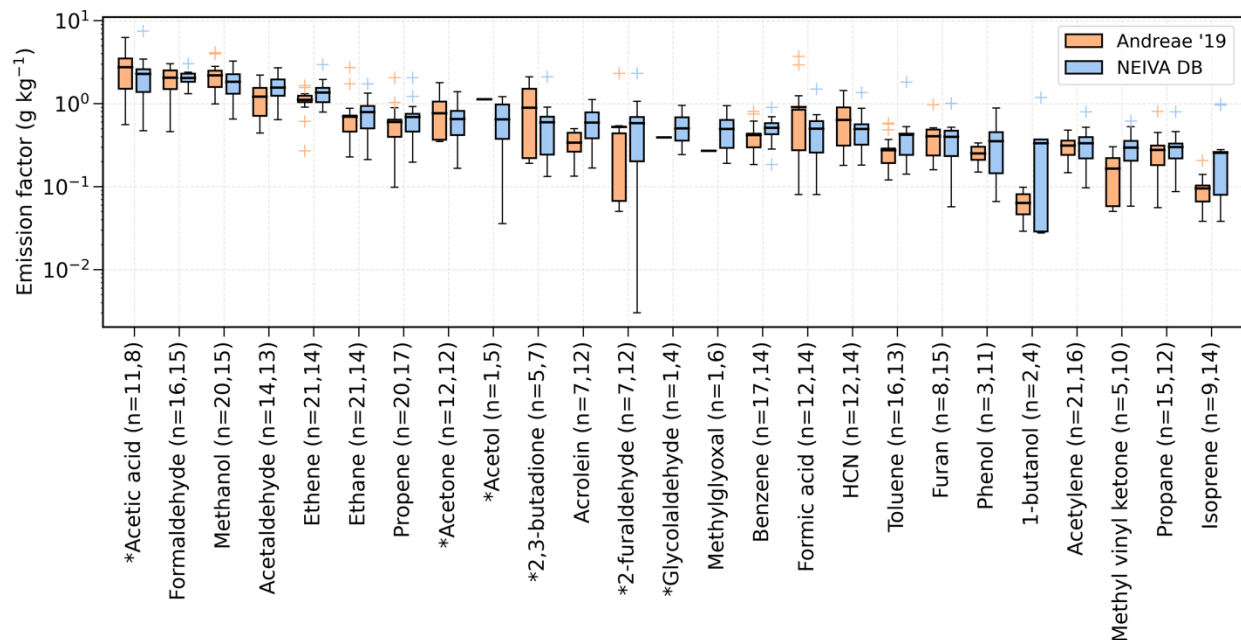


Figure S6: The 25 most abundant NMOC_g EFs for temperate forest. The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers ($> 1.5 \times$ above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“n”) with Andreae followed by NEIVA.

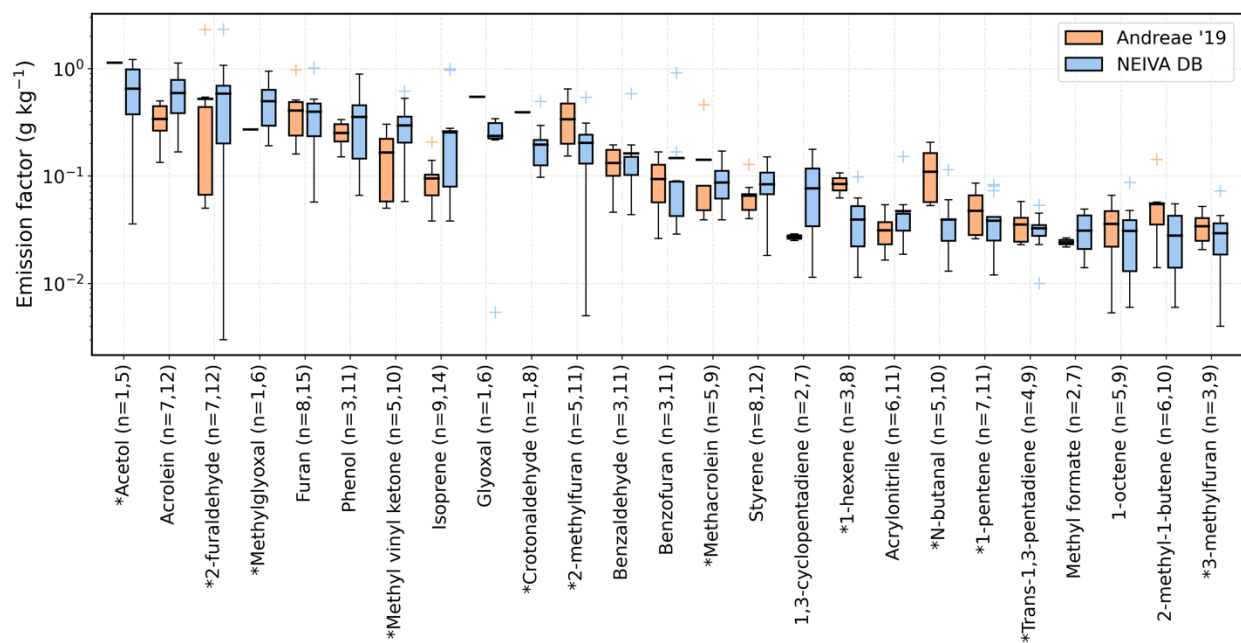


Figure S7: NMOC_g EFs for temperate forest in which n NEIVA > n Andreae (where “n” is the number of observations). The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers (> 1.5 × above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“n”) with Andreae followed by NEIVA.

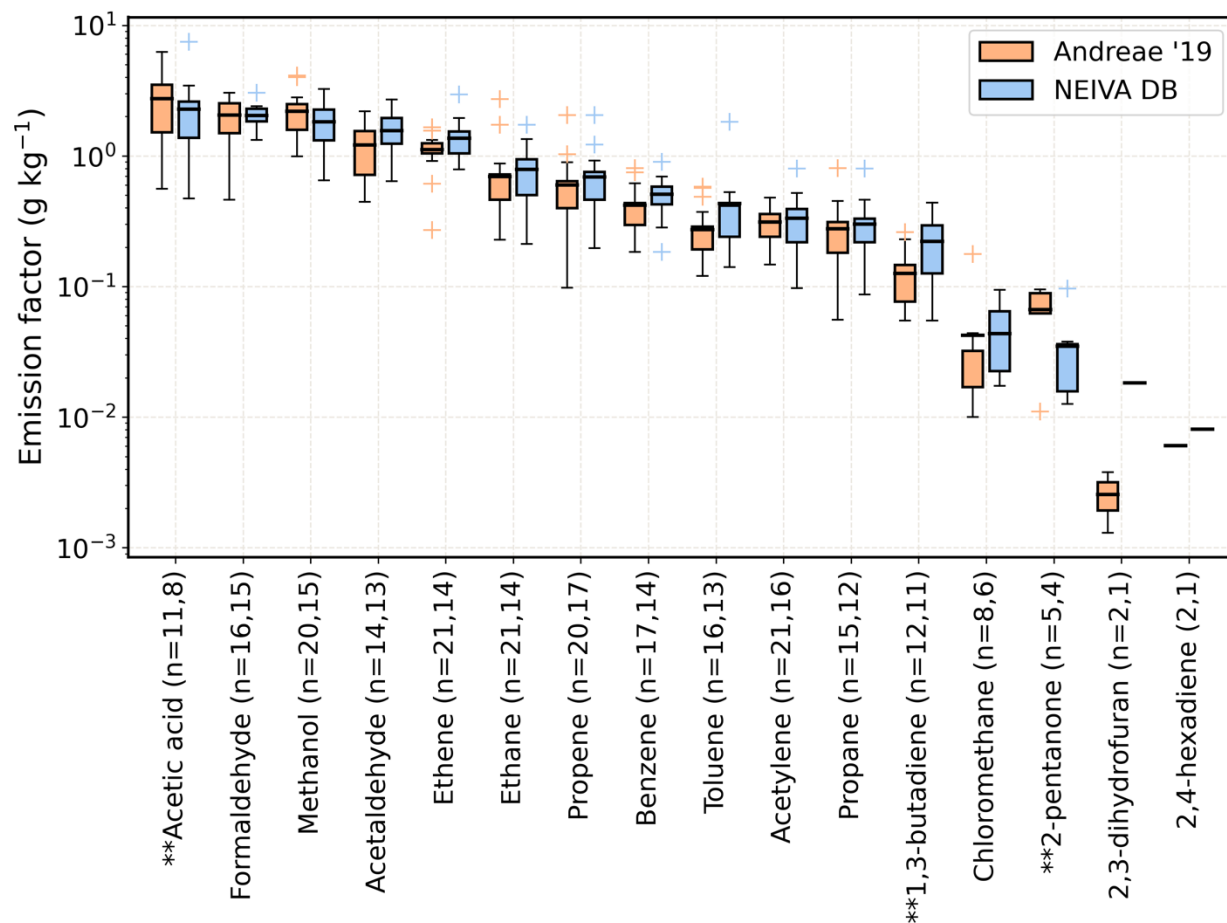


Figure S8: NMOG_g EFs for temperate forest in which $n_{\text{NEIVA}} < n_{\text{Andreae}}$ (where “n” is the number of observations). The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers ($> 1.5 \times$ above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“n”) with Andreae followed by NEIVA.

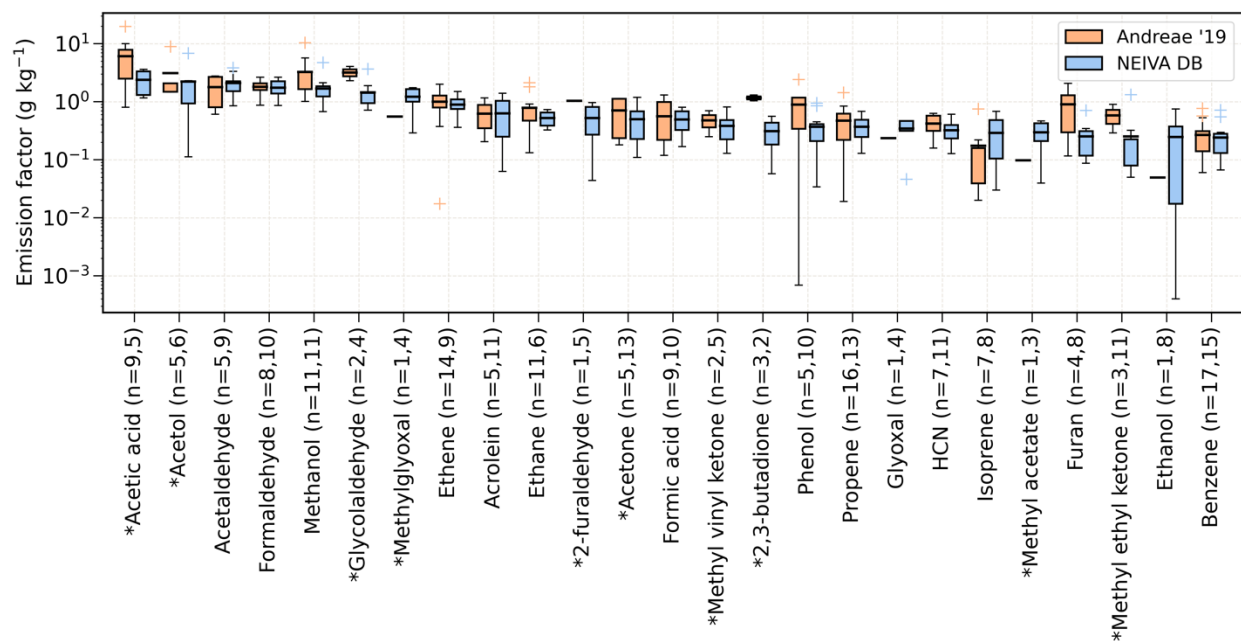


Figure S9: The 25 most abundant NMOC_g EFs for crop residue. The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers ($> 1.5 \times$ above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“n”) with Andreae followed by NEIVA.

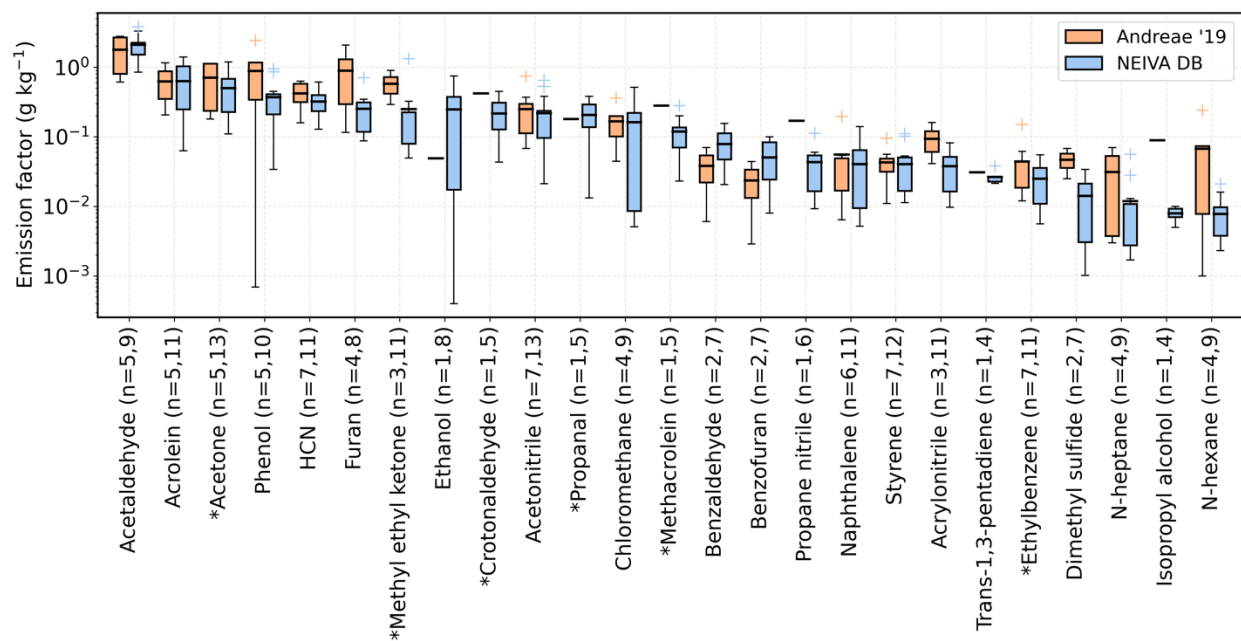


Figure S10: NMOC_g EFs for crop residue in which $n_{\text{NEIVA}} > n_{\text{Andreae}}$ (where “n” is the number of observations). The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers ($> 1.5 \times$ above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“n”) with Andreae followed by NEIVA.

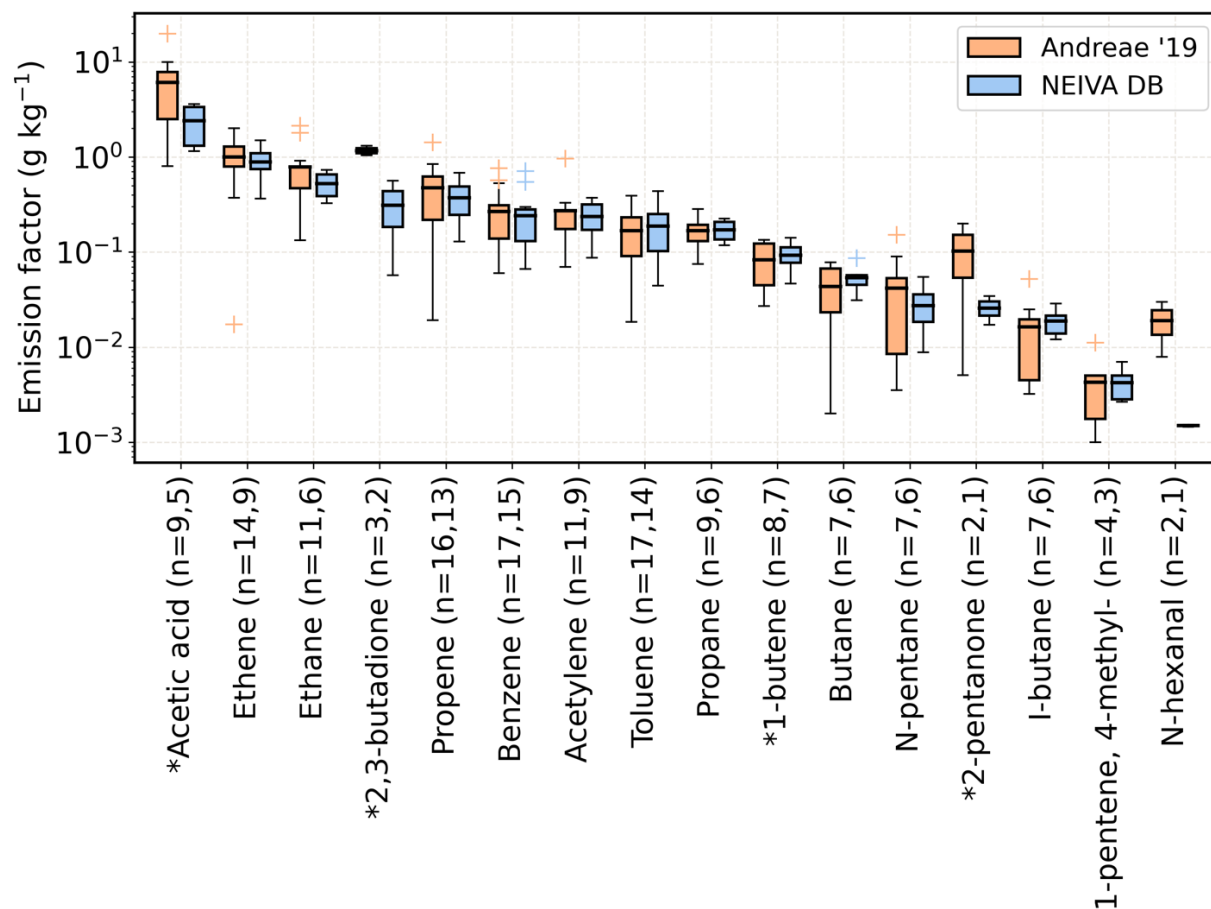


Figure S11: NMOG_g EFs for crop residue in which n NEIVA < n Andreae (where “ n ” is the number of observations). The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers ($> 1.5 \times$ above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“ n ”) with Andreae followed by NEIVA.

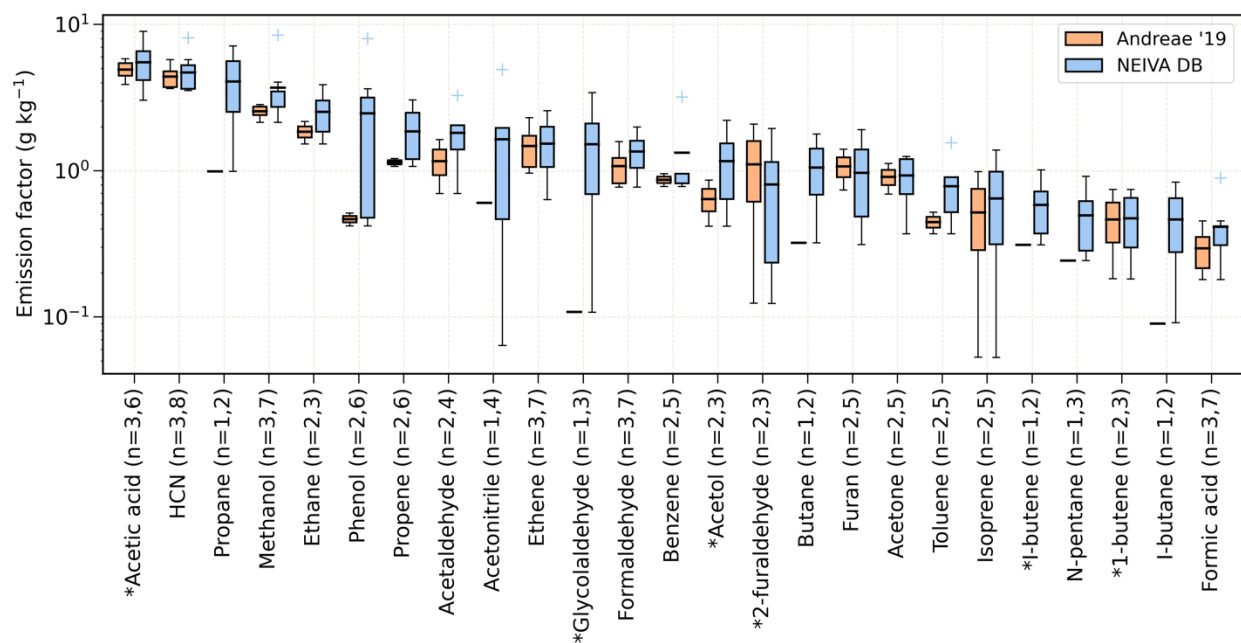


Figure S12: The 25 most abundant NMOC_g EFs for peat. The black lines indicate the mean values and are equivalent to the recommended EFs in the NEIVA database. The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers ($> 1.5 \times$ above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“n”) with Andreae followed by NEIVA.

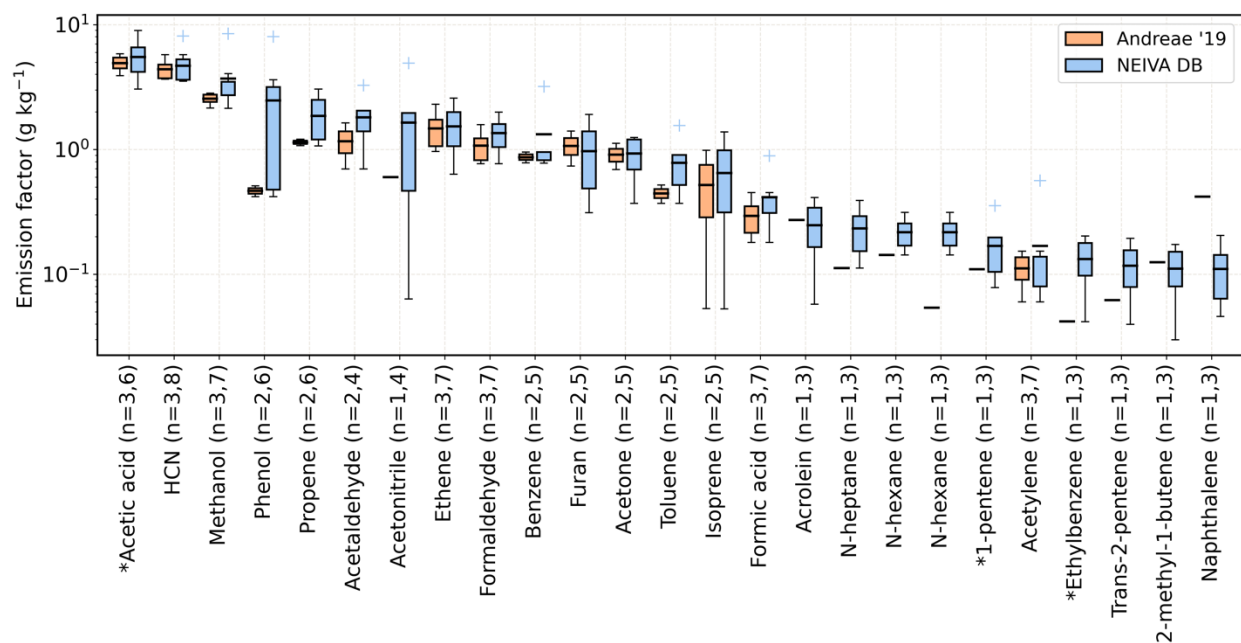


Figure S13: NMOC_g EFs for peat in which n NEIVA > n Andreae (where “n” is the number of observations). The box and whiskers represent the data in the processed EF dataset. The black lines indicate the mean values and are equivalent to the recommended EFs. The outliers ($> 1.5 \times$ above/below the interquartile range) are indicated by the plus symbols. Compounds marked with an asterisk before the name have had application of isomeric distributions. The number of observations is listed in parenthesis (“n”) with Andreae followed by NEIVA.

Numbers of Compounds Needed to Represent 90% of the Total NMOC_g EF

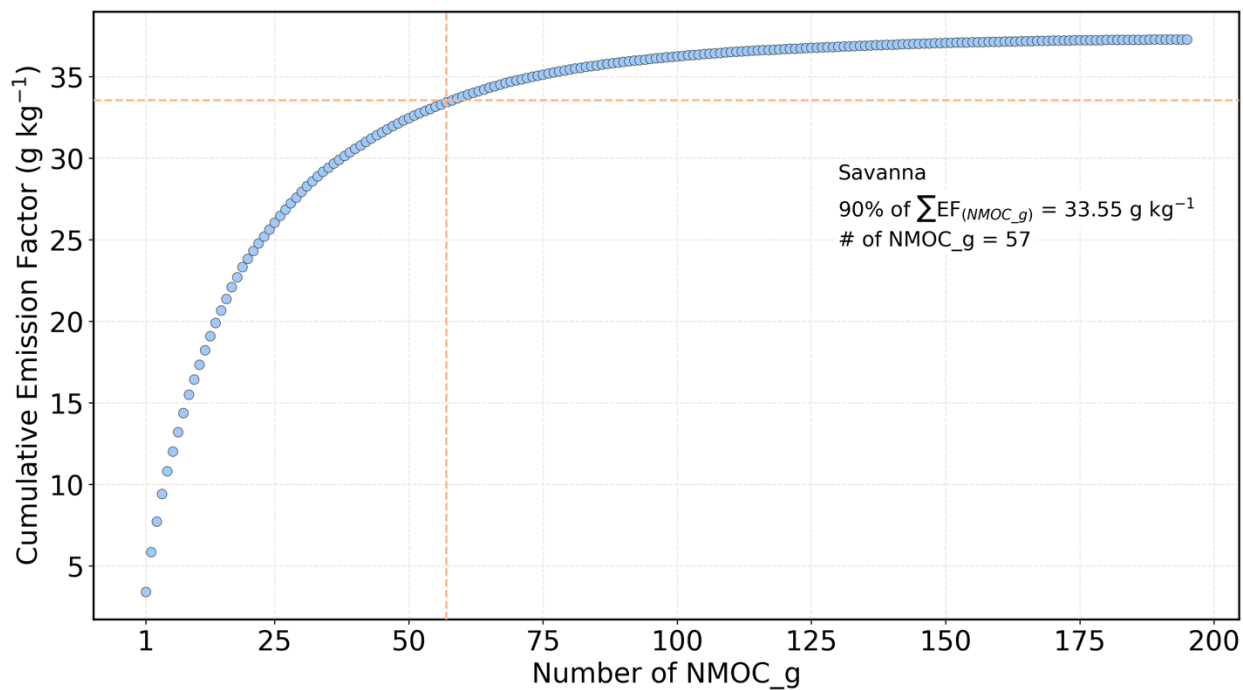


Figure S14: Number of compounds needed to represent 90% savanna NMOC_g EF.

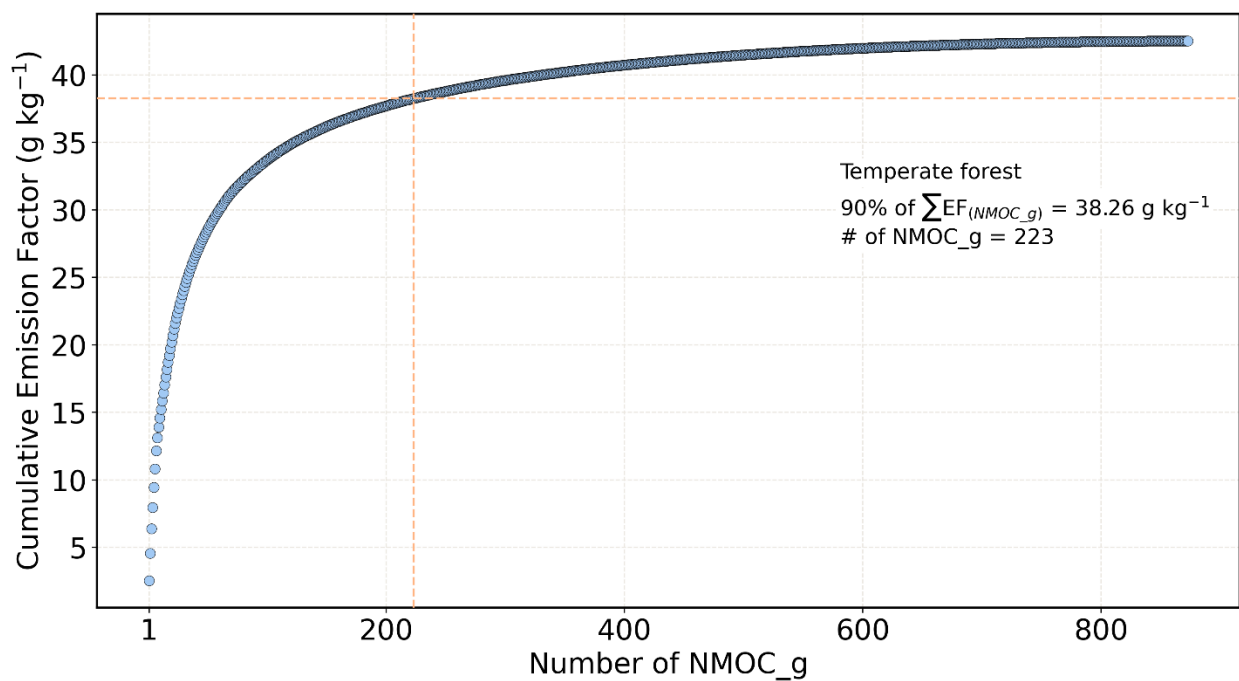


Figure S15: Number of compounds needed to represent 90% temperate forest NMOC_g EF.

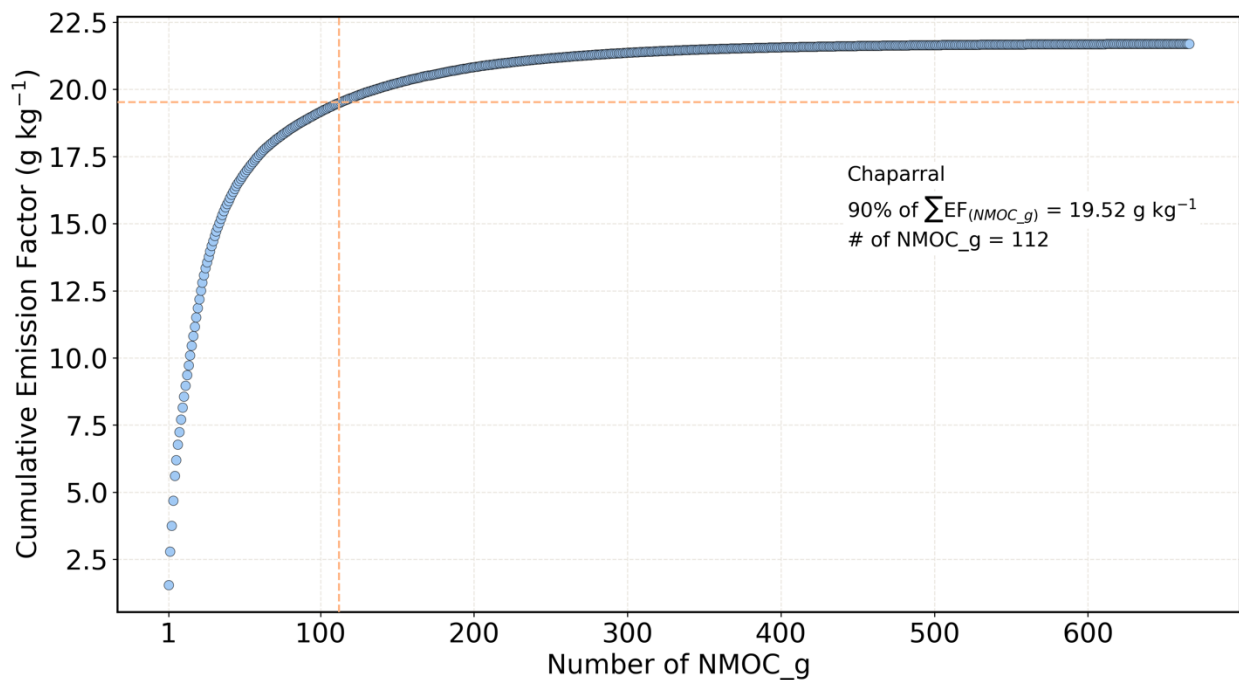


Figure S16: Number of compounds needed to represent 90% chaparral NMOC_g EF.

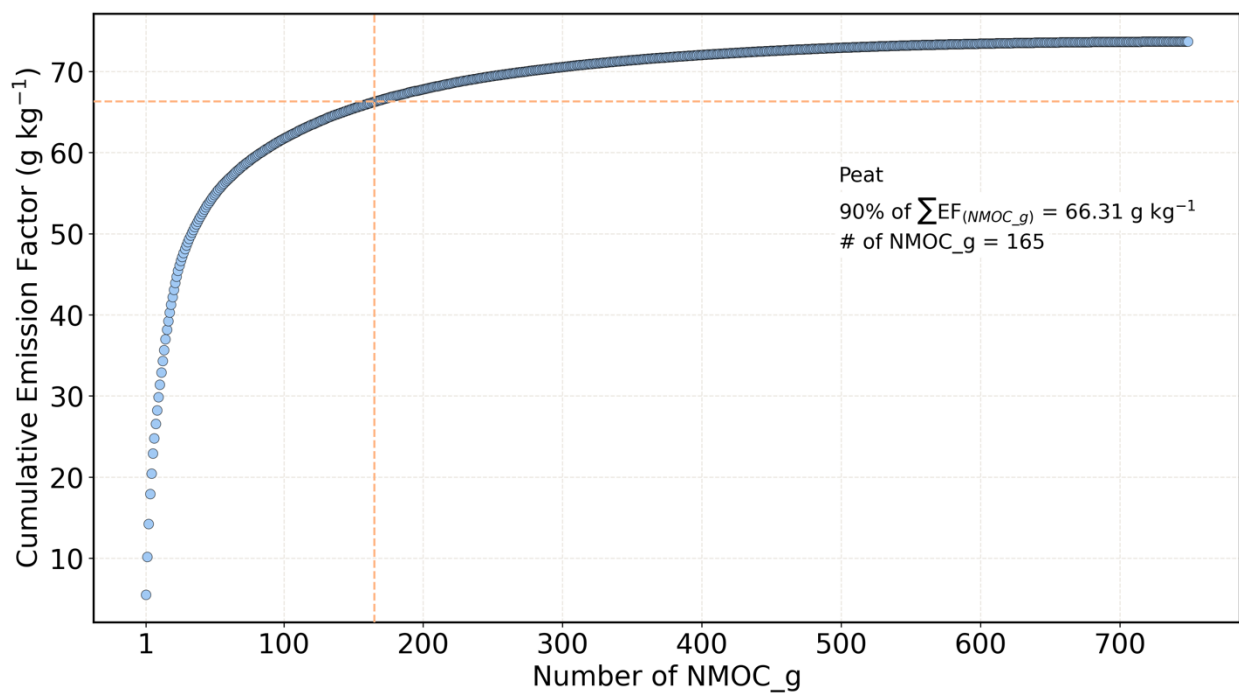


Figure S17: Number of compounds needed to represent 90% peat NMOC_g EF.

S8. Backend Database

This database contains tables that are used within the python script files to perform the data processing and data integration steps, to produce the datasets, and to query the datasets.

Table S23: Description tables comprises the backend database.		
Category	Table Name	Description
Information tables for data processing and query functions.	bkdb_info_efcol_integrated_ef	This information table contains the EF column names from the integrated EF dataset, along with related details such as fire type, measurement type, reference study, and publication year.
	bkdb_info_efcol_processed_ef	This information table contains the EF column names from the processed EF dataset, along with related details such as fire type, measurement type, reference study, and publication year.
	bkdb_info_table_name	This table includes the names of tables from the legacy, raw, and primary databases, along with details like fire type, measurement type, reference publication (DOI), year of publication, and reference to the source table.
	bkdb_info_efcol_rdb_ldb	This table include the EF column names of from all tables in the legacy and raw database, along with related information such as fire type, measurement type.
Data processing table. Used in data processing from raw database to primary database.	bkdb_correction_factor	For the % C correction described in S2, this dataset includes the EF column names and associated % C correction factor.
These tables are produced during the data integration process, and used for merging compounds that do not have an InChI.	bkdb_nmog_LumCom	This dataset is a subset of the integrated dataset and includes lumped compounds with a '+' sign in the compound name extracted from the integrated dataset.
	bkdb_nmog_LumCom_altName	This dataset is a replicate of 'bkdb_nmog_LumCom,' with the inclusion of an extra column named 'altered_name.' This column provides users with the flexibility to adjust the lumped compound name(s) or constituent compound(s) name(s).
	bkdb_nmog_MultLumCom	This dataset contains lumped compounds that have more than one group of lumped compounds for a given formula.
	bkdb_nmog_MultLumCom_slc_id	In this dataset lumped compounds with the same chemical formula are consolidated into a single row (i.e., considered the same compound). The representative compound name is selected based on the longest name among the lumped compounds and the corresponding id is selected.
	bkdb_nmog_MultLumCom_slc_id_altName	This dataset is the same as the 'bkdb_nmog_MultLumCom_slc_id,' but it includes an extra column named 'altered_name,' This column provides users

		with the flexibility to adjust the lumped compound name(s) or constituent compound(s) name(s).
Used for data sorting following the data integration step.	bkdb_pm_order_seq	This dataset lists the PM constituents in the following order: PM size, PM organic, PM elemental, PM ion, PM metal and NMOC_p.
Used during the ER adjustment data processing step.	bkdb_compound_flaming_combustion	The list of flaming-dominant compounds and their unique ids.
These datasets are produced during data integration and used during the fractional contribution data processing step.	bkdb_fc_calc_simple	The list of lumped compounds without specified compounds.
	bkdb_fc_calc_specific	The list of lumped compounds with specified compounds.
Chemical mechanism and property assignment tables. These tables are used in generating the property and reactivity profiles.	chem_property_h15isomers	This information table includes compounds reported by Hatch et al. (2015) with Type-2 IDs, which represent isomeric compounds with unknown chemical structures, that were assigned model surrogate and properties based on the nearest identified compound. This proximity was determined using the first- and second-dimension retention indices.
	chem_property_inchi	This information table includes compounds with Type 1 IDs, which represent single identified compounds, that were assigned model surrogate and properties based on SMILES and functional group type(s).
	chem_property_lumpCom	This information table includes lumped compounds without specified constituent compounds, that were assigned model surrogate and properties based on the descriptive names, such as C11 Aromatics and C9 Nitriles.
	chem_property_lumpCom_spec	This information table includes lumped compounds with specified constituent compounds, that were assigned model surrogate and properties based on their individual components.

Table S24: Description of functions that display information.		
Function name	Input	Description
fire_type()	Not required.	This function returns the list of fire types.
table_info()	Database, fire type	This function returns a list of table names along with associated information such as measurement type, publication DOI, pollutant category for a specified database name and fire type.
summary_table()	Fire type, measurement type	This function returns a list of table names along with associated information such as measurement type, publication DOI, pollutant category for a specified database name and fire type.
display_pollutant_category()	Not required.	This function displays the list of pollutant category of recommended EF table.
property_variables()		This function provides descriptions of property variables, their units, and the data sources associated with them .The Property_Surrogate table comprises atmospheric modeling-relevant property variables of compounds. The chemical mechanism surrogate and property dataset is discussed in section S6.
model_surrogate()	Chemical mechanism (e.g., S07, S07T, S18B, S22, MOZT1, GEOSChem)	This function retrieves the unique set of model surrogates for a specified chemical mechanism.

Table S25: Description of functions used to query EF data and properties of NMOC_g.		
Function name	Input	Description
<code>select_compound()</code>	Fire type, compound, table name	The function returns the EF data for the specified fire type, compound name, and table name. If the specified compound is not found, the function returns the statement 'Compound not found. Search by formula'. For example, 'Eugenol + isoeugenol' is a lumped compound. Therefore, searching for 'eugenol' or 'isoeugenol' individually may not yield results. However, using the <code>select_chemical_formula()</code> function allows users to search by formula, where both lumped and individual compounds will appear.
<code>select_chemical_formula()</code>	Fire type, formula, table name	This function returns the emission factor data for a specified fire type, chemical formula, and table name.
<code>select_pm_data()</code>	Fire type, table name	This function returns the EFs for the PM constituents for the specified fire type and table name.
<code>select_ef_pollutant_category()</code>	Fire type, pollutant category	The function returns the recommended EFs for the specified fire type and pollutant category. The pollutant category options can be obtained using the <code>pollutant_category()</code> function.
<code>compare_lab_field()</code>	Fire type, compound, table name	This function returns mean EF, mean MCE, and number of observations of laboratory versus field data for the specified fire type, compound name, and table name.
<code>ef_sorted_by_property()</code>	Fire type, chemical mechanism, model surrogate, property variable	This function returns the NMOC_g EFs sorted by the specified property variable in ascending order. The NMOC_g are filtered by the specified fire type, chemical mechanism, and model surrogate.
<code>plot_ef()</code>	Fire type, compound, table name	This function generates a plot illustrating the recommended EF for a specified fire type, compound, and table name.
<code>boxplot_ef()</code>	Fire type, compound, table name	This function is identical to <code>plot_ef()</code> but it generates a boxplot of the EF data instead of a single point.
<code>mce_vs_ef()</code>	Compound, fire type	This function generates a scatter plot of MCE versus EF and returns the linear fit coefficients if the data count is greater than 4 for a specified compound and fire type.
<code>voc_profile()</code>	Fire type, chemical mechanism	The function returns the mole fraction of the model surrogates of the specified chemical mechanism and fire type.

<code>calc_ohr()</code>	Fire type, chemical mechanism, summation of NMOC_g EF in ppb	Calculates the OHR based a specified chemical mechanism and fire type. The function requires the summation of NMOC_g EF in ppb as an input for the calculation step, which are then mapped to the model surrogates based on mole fraction.
<code>calc_vbs()</code>	Fire type	Calculates and displays the volatility basis set for the specified fire type.
<code>weighted_property()</code>	Fire type, chemical mechanism	Calculates the molecular weight (mm), OH rate constant (kOH), logarithm of saturation concentration (cstar), vapor pressure (vp), and Henry's law constant (hc) weighted by the EF of each model surrogate for the specified chemical mechanism and fire type.
<code>speciation_profile()</code>	Fire type, chemical mechanism, model surrogate	This function returns the NMOC_g EFs for a specified chemical mechanism and fire type.
<code>abundant_nmog()</code>	Fire type, chemical mechanism, property variable	The function returns the 25 most abundant NMOC_g sorted by EF in ascending order for a specified fire type. The input chemical mechanism model surrogate and property variable will be mapped to the 25 compounds and displayed in the output table.
<code>boxplot_abundant_nmog()</code>	Fire type	Generates a boxplot displaying the 25 most abundant NMOC_g compounds sorted by EF in ascending order for a specified fire type.
<code>nmog_with_high_n()</code>	Fire type, chemical mechanism, property variable	This function is identical as the <code>abundant_nmog()</code> function, but it outputs the 25 NMOC_g sorted by data count in ascending order.
<code>nmog_with_high_ohr()</code>	Fire type, chemical mechanism, summation of NMOC_g EF in ppb	The function returns the 25 NMOC_g sorted by OH reactivity (OHR) in ascending order for a specified fire type, chemical mechanism, and summation of all NMOC_g in ppb unit. The input chemical mechanism model surrogate will be mapped to the 25 compounds.
<code>plot_model_surrogate()</code>	Fire type, chemical mechanism, model surrogate	Generates a boxplot illustrating NMOC_g versus EF. The NMOC_g compounds are sorted by EF in ascending order and are mapped to the specified model surrogate and chemical mechanism. If there are more than 25 compounds under the model surrogate of the chemical mechanism, it plots the first 25 compounds.

The following function was demonstrated in the main manuscript. Some of the resultant tables were moved here due to their length.

`select_pm_data(fire_type, table_name)`. This function returns the EFs in all PM subcategories (e.g., PM size, PM organic, PM elemental, PM ion, PM metal, NMOC_p and PM optical property) for the specified fire type. The tables below are separated by subcategory. The other subcategories are presented in the main manuscript.

using inputs (peat, integrated EF):

pollutant_category-PM metal					
mm	formula	compound	EF_peat_jayarathne18	EF_russia_watson19	EF_siberia_watson19
58.69	Ni	Nickel			3.00E-04
63.55	Cu	Copper	1.28E-02	2.30E-03	8.40E-03
74.92	As	Arsenic	1.00E-04	2.00E-04	1.00E-04
78.97	Se	Selenium		2.00E-04	4.00E-04
112.41	Cd	Cadmium		1.10E-03	5.00E-04
121.76	Sb	Antimony			
207.00	Pb	Lead	7.00E-04	1.10E-03	8.00E-04
55.84	Fe	Fe	4.70E-03	2.23E-02	8.60E-03
65.40	Zn	Zn	6.90E-03	1.40E-03	1.70E-03
137.33	Ba	Ba	2.00E-04	5.10E-03	3.90E-03
26.98	Al	Aluminum		6.27E-02	2.42E-02
28.09	Si	Silicon		4.20E-03	4.70E-03
30.97	P	Phosphorus		1.00E-04	
32.07	S	Sulfur		1.48E-02	1.86E-02
44.96	Sc	Scandium		1.36E-02	1.79E-02
47.87	Ti	Titanium		2.20E-03	1.40E-03
50.94	V	Vanadium			1.00E-04
52.00	Cr	Chromium		3.00E-04	3.00E-04
54.94	Mn	Manganese		1.00E-03	9.00E-04
159.81	Br	Bromine		3.00E-04	2.60E-03

85.47	Rb	Rubidium		6.00E-04	2.00E-04
87.60	Sr	Strontium		1.40E-03	1.50E-03
88.91	Y	Yttrium		9.00E-04	1.00E-03
91.22	Zr	Zirconium		1.10E-03	1.80E-03
92.91	Nb	Niobium		2.00E-04	1.00E-04
96.00	Mo	Molybdenum		8.00E-04	8.00E-04
107.87	Ag	Silver		2.00E-04	1.10E-03
114.82	In	Indium		1.20E-03	5.00E-04
118.71	Sn	Tin		9.00E-04	3.00E-03
132.91	Cs	Cesium		9.70E-03	2.10E-03
138.91	La	Lanthanum		1.97E-02	1.62E-02
183.80	W	Wolfram		1.80E-03	8.00E-04
196.97	Au	Gold		8.00E-04	5.00E-04
238.03	U	Uranium		6.00E-04	1.00E-03

pollutant_category-PM metal (continued)					
mm	formula	compound	EF_northern_alaska_watson19	EF_evergladesNP_florida_watson19	EF_malaysia_watson19
58.69	Ni	Nickel	1.00E-04		4.00E-04
63.55	Cu	Copper	4.90E-03	1.60E-03	1.60E-03
74.92	As	Arsenic	1.00E-04	1.00E-04	
78.97	Se	Selenium	2.00E-04	6.00E-04	3.00E-04
112.41	Cd	Cadmium		1.00E-03	
121.76	Sb	Antimony	6.00E-04	1.00E-04	1.20E-03
207.00	Pb	Lead		5.00E-04	5.00E-04
55.84	Fe	Fe	8.30E-03	4.60E-03	1.36E-02
65.40	Zn	Zn	3.60E-03	7.00E-04	8.00E-04
137.33	Ba	Ba			
26.98	Al	Aluminum	5.60E-03	3.50E-03	9.40E-03

28.09	Si	Silicon	1.30E-03	2.20E-03	1.10E-03
30.97	P	Phosphorus			
32.07	S	Sulfur	4.20E-03	9.90E-02	1.43E-02
44.96	Sc	Scandium			
47.87	Ti	Titanium	8.00E-04	8.00E-04	1.50E-03
50.94	V	Vanadium		1.00E-04	
52.00	Cr	Chromium	1.00E-04		3.00E-04
54.94	Mn	Manganese	5.00E-04	5.00E-04	1.60E-03
159.81	Br	Bromine	1.10E-03	5.90E-03	2.50E-03
85.47	Rb	Rubidium	5.00E-04	2.00E-04	1.00E-04
87.60	Sr	Strontium	9.00E-04	1.30E-03	4.00E-04
88.91	Y	Yttrium	7.00E-04	7.00E-04	4.00E-04
91.22	Zr	Zirconium	7.00E-04	9.00E-04	6.00E-04
92.91	Nb	Niobium		5.00E-04	4.00E-04
96.00	Mo	Molybdenum	3.00E-04	4.00E-04	6.00E-04
107.87	Ag	Silver		2.00E-04	6.00E-04
114.82	In	Indium	2.00E-04	1.50E-03	5.00E-04
118.71	Sn	Tin	1.50E-03	1.70E-03	7.00E-04
132.91	Cs	Cesium	1.80E-03	4.90E-03	3.60E-03
138.91	La	Lanthanum	8.50E-03	1.10E-02	6.00E-03
183.80	W	Wolfram	5.00E-04	2.30E-03	
196.97	Au	Gold		6.00E-04	1.00E-04
238.03	U	Uranium	9.00E-04	8.00E-04	6.00E-04

pollutant_category- NMOC_p					
mm	formula	compound	EF_northern_ alaska watson19	EF_evergladesNP_ florida watson19	EF_malaysia_ watson19
178.23	C14H10	Anthracene	1.00E-04		
202.25	C16H10	Fluoranthene	4.00E-04		
202.25	C16H10	Pyrene	7.00E-04		
216.28	C17H12	Methylfluoranthene	5.00E-04		
192.25	C15H12	9-Methylanthracene			
226.30	C18H10	Benzo(ghi)fluoranthene	1.00E-04		
226.30	C18H10	Cyclopenta(cd)pyrene	1.00E-04		
228.30	C18H12	Benz(a)anthracene	3.00E-04		
228.30	C18H12	Chrysene	7.00E-04		
242.30	C19H14	1-Methylchrysene	2.00E-04		
234.30	C18H18	Retene	4.00E-04		
252.30	C20H12	Benzo(b)fluoranthene	3.00E-04		
252.30	C20H12	Benzo(k)fluoranthene			
252.30	C20H12	Benzo(j)fluoranthene			
252.30	C20H12	Benzo(e)pyrene	4.00E-04		
252.30	C20H12	Benzo(a)pyrene	1.00E-04		
252.30	C20H12	Perylene	1.00E-04		
276.30	C22H12	Indeno(1,2,3-cd)pyrene			
276.30	C22H12	Benzo(ghi)perylene	2.00E-04		
278.30	C22H14	Dibenz[a,h]anthracene	1.00E-04		
278.30	C22H14	Picene	2.00E-04		
306.40	C24H18	Triphenylbenzene			
370.65	C27H46	17(H)-22,29,30-Trisnorhopane	4.30E-03		
268.50	C19H40	Pristane	1.24E-02		

254.50	C18H38	Norpristane	4.30E-03		
282.50	C20H42	Phytane			
422.80	C30H62	Squalane	1.62E-02		
254.50	C18H38	Octadecane	4.80E-03		
268.50	C19H40	Nonadecane	1.36E-02		
282.50	C20H42	Eicosane	2.73E-02		
296.60	C21H44	Heneicosane	4.71E-02		
310.60	C22H46	Docosane	5.33E-02		
324.60	C23H48	Tricosane	5.95E-02		
338.70	C24H50	Tetracosane	5.08E-02		
352.70	C25H52	Pentacosane	6.70E-02		
366.70	C26H54	Hexacosane	5.08E-02		
380.70	C27H56	Heptacosane	6.82E-02		
394.80	C28H58	Octacosane	5.95E-02		
408.80	C29H60	Nonacosane	8.06E-02		
422.80	C30H62	triacontane	5.83E-02		
436.80	C31H64	Hentriacontane	8.31E-02		
450.90	C32H66	Dotriacontane	3.76E-02		
464.90	C33H68	Tritriacontane	3.51E-02		
478.90	C34H70	Tetratriacontane	1.55E-02		
492.90	C35H72	Pentatriacontane	8.20E-03		
162.14	C6H10O5	Levogluconan	5.70E-01	1.58E+01	2.63E+00
521.00	C37H76	Heptatriacontane	1.02E-02		
535.00	C38H78	Octriacontane	3.10E-02		
549.10	C39H80	Nonatriacontane	1.22E-02		
162.14	C6H10O5	Mannosan	1.15E-02	1.93E+00	2.06E-01
162.14	C6H10O5	Galactosan	1.70E-03		

180.16	C6H12O6/C 12H24O11	Galactose/Maltitol		1.58E-02	1.10E-03
92.09	C3H8O3	Glycerol		1.21E+00	7.85E-02
182.17	C6H14O6	Mannitol		1.58E-02	
182.17	C9H10O4	Syringaldehyde	1.15E-02		
168.15	C8H8O4	Vanillic acid	4.59E-02		
198.17	C9H10O5	Syringic acid	2.10E-02		
386.70	C27H46O	Cholesterol			
412.70	C29H48O	Stigmasterol	2.70E-03		
414.70	C29H50O	b-Sitosterol	6.60E-03		
400.70	C28H48O	Campesterol	3.60E-03		
388.67	C27H48O	Cholestanol and coprostanol			
416.70	C29H52O	Stigmastanol			
370.65	C27H46	17(H)-22,29,30- Trisnorhopane			
398.71	C29H50	17(H)-21(H)-30- Norhopane	1.05E-02		
412.73	C30H52	17(H)-21(H)-Hopane	2.70E-03		
178.23	C14H10	Anthracene	1.00E-04		

pollutant_category- NMOC_p (continued)					
mm	formula	compound	EF_northern_ alaska_watson19	EF_evergladesNP_ florida_watson19	EF_malaysia_ watson19
178.23	C14H10	Anthracene			
202.25	C16H10	Fluoranthene			
202.25	C16H10	Pyrene			
216.28	C17H12	Methylfluoranthene			
192.25	C15H12	9-Methylanthracene			
226.30	C18H10	Benzo(ghi)fluoranthene			

226.30	C18H10	Cyclopenta(cd)pyrene			
228.30	C18H12	Benz(a)anthracene			
228.30	C18H12	Chrysene			
242.30	C19H14	1-Methylchrysene			
234.30	C18H18	Retene			
252.30	C20H12	Benzo(b)fluoranthene			
252.30	C20H12	Benzo(k)fluoranthene			
252.30	C20H12	Benzo(j)fluoranthene			
252.30	C20H12	Benzo(e)pyrene			
252.30	C20H12	Benzo(a)pyrene			
252.30	C20H12	Perylene			
276.30	C22H12	Indeno(1,2,3-cd)pyrene			
276.30	C22H12	Benzo(ghi)perylene			
278.30	C22H14	Dibenz[a,h]anthracene			
278.30	C22H14	Picene			
306.40	C24H18	Triphenylbenzene			
370.65	C27H46	17(H)-22,29,30-Trisnorhopane			
268.50	C19H40	Pristane			
254.50	C18H38	Norpristane			
282.50	C20H42	Phytane			
422.80	C30H62	Squalane			
254.50	C18H38	Octadecane			
268.50	C19H40	Nonadecane			
282.50	C20H42	Eicosane			
296.60	C21H44	Heneicosane			
310.60	C22H46	Docosane			
324.60	C23H48	Tricosane			

338.70	C24H50	Tetracosane			
352.70	C25H52	Pentacosane			
366.70	C26H54	Hexacosane			
380.70	C27H56	Heptacosane			
394.80	C28H58	Octacosane			
408.80	C29H60	Nonacosane			
422.80	C30H62	Triacontane			
436.80	C31H64	Hentriacontane			
450.90	C32H66	Dotriacontane			
464.90	C33H68	Tritriacontane			
478.90	C34H70	Tetratriacontane			
492.90	C35H72	Pentatriacontane			
162.14	C6H10O5	Levogluconan	3.42E+00	3.93E-01	7.81E-01
521.00	C37H76	Heptatriacontane			
535.00	C38H78	Octriacontane			
549.10	C39H80	Nonatriacontane			
162.14	C6H10O5	Mannosan	6.94E-01	7.40E-03	2.04E-02
162.14	C6H10O5	Galactosan			
180.16	C6H12O6/C 12H24O11	Galactose/Maltitol			
92.09	C3H8O3	Glycerol	1.82E-01		
182.17	C6H14O6	Mannitol			1.50E-03
182.17	C9H10O4	Syringaldehyde			
168.15	C8H8O4	Vanillic acid			
198.17	C9H10O5	Syringic acid			
386.70	C27H46O	Cholesterol			
412.70	C29H48O	Stigmasterol			
414.70	C29H50O	b-Sitosterol			

400.70	C28H48O	Campesterol			
388.67	C27H48O	Cholesterol and coprostanol			
416.70	C29H52O	Stigmastanol			
370.65	C27H46	17(H)-22,29,30-Trisnorhopane			
398.71	C29H50	17(H)-21(H)-30-Norhopane			
412.73	C30H52	17(H)-21(H)-Hopane			
178.23	C14H10	Anthracene			

References

- Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D., and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, *Atmospheric Chemistry and Physics*, 11, 4039–4072, <https://doi.org/10.5194/acp-11-4039-2011>, 2011.
- Andreae, M. O.: Emission of trace gases and aerosols from biomass burning – an updated assessment, *Atmospheric Chemistry and Physics*, 19, 8523–8546, <https://doi.org/10.5194/acp-19-8523-2019>, 2019.
- Bertschi, I. T., Yokelson, R. J., Ward, D. E., Christian, T. J., and Hao, W. M.: Trace gas emissions from the production and use of domestic biofuels in Zambia measured by open-path Fourier transform infrared spectroscopy, *Journal of Geophysical Research: Atmospheres*, 108, <https://doi.org/10.1029/2002JD002158>, 2003.
- Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *Journal of Geophysical Research: Atmospheres*, 106, 23073–23095, <https://doi.org/10.1029/2001JD000807>, 2001.
- Burling, I., Yokelson, R., Akagi, S., Urbanski, S., Wold, C., Griffith, D., Johnson, T., Reardon, J., and Weise, D.: Airborne and ground-based measurements of the trace gases and particles emitted by prescribed fires in the United States, *Atmospheric Chemistry and Physics*, 11, 12197–12216, <https://doi.org/10.5194/acp-11-12197-2011>, 2011.
- Carter, T. S., Heald, C. L., Kroll, J. H., Apel, E. C., Blake, D., Coggon, M., Edtbauer, A., Gkatzelis, G., Hornbrook, R. S., Peischl, J., Pfannerstill, E. Y., Piel, F., Reijrink, N. G., Ringsdorf, A., Warneke, C., Williams, J., Wisthaler, A., and Xu, L.: An improved representation of fire non-methane organic gases (NMOGs) in models: emissions to reactivity, *Atmospheric Chemistry and Physics*, 22, 12093–12111, <https://doi.org/10.5194/acp-22-12093-2022>, 2022.
- Carter, W. P. L.: Development of the SAPRC-07 Chemical Mechanism, *Atmospheric Environment*, 44, 5324–5335, <https://doi.org/10.1016/j.atmosenv.2010.01.026>, 2010.
- Carter, W. P. L.: Documentation of the SAPRC-16 Mechanism Generation System, Draft interim report to California Air Resources Board Contract 11-761. May 25. Available at <https://intra.engr.ucr.edu/~carter/SAPRC/18/MechGen.pdf>, 2019.
- Carter, W. P. L.: Documentation of the SAPRC-18 Mechanism, Report to California Air Resources Board Contract No. 11-761, May 29. Available at <https://intra.cert.ucr.edu/~carter/SAPRC/18/S18doc.pdf>, 2020.
- Carter, W. P. L.: Documentation of the SAPRC-22 Mechanisms, Report to the California Air Resources Board contract no. 21AQP011, May 15. Available at <https://intra.cert.ucr.edu/~carter/SAPRC/22/S22doc.pdf>, 2023a.

Carter, W. P. L.: Development of an Improved Chemical Speciation Database for Processing Emissions of Volatile Organic Compounds for Air Quality Models, Available at <https://intra.engr.ucr.edu/~carter/emitdb/>, 2023b.

Carter, W. P. L.: The SAPRC Mechanism Generation System, Available at <https://intra.cert.ucr.edu/~carter/SAPRC/22/MechGen.pdf>, In Prep.

Christian, T., Kleiss, B., Yokelson, R., Holzinger, R., Crutzen, P., Hao, W., Saharjo, B., and Ward, D.: Comprehensive laboratory measurements of biomass-burning emissions: 1. Emissions from Indonesian, African, and other fuels, *Journal of Geophysical Research: Atmospheres*, 108, <https://doi.org/10.1029/2003JD003704>, 2003.

Emmons, L. K., Schwantes, R. H., Orlando, J. J., Tyndall, G., Kinnison, D., Lamarque, J.-F., Marsh, D., Mills, M. J., Tilmes, S., Bardeen, C., Buchholz, R. R., Conley, A., Gettelman, A., Garcia, R., Simpson, I., Blake, D. R., Meinardi, S., and Pétron, G.: The Chemistry Mechanism in the Community Earth System Model Version 2 (CESM2), *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001882, <https://doi.org/10.1029/2019MS001882>, 2020.

Hatch, L. E., Luo, W., Pankow, J. F., Yokelson, R. J., Stockwell, C. E., and Barsanti, K. C.: Identification and quantification of gaseous organic compounds emitted from biomass burning using two-dimensional gas chromatography-time-of-flight mass spectrometry, *Atmospheric Chemistry and Physics*, 15, 1865–1899, <https://doi.org/10.5194/acp-15-1865-2015>, 2015.

Hutzell, W. T., Luecken, D. J., Appel, K. W., and Carter, W. P. L.: Interpreting predictions from the SAPRC07 mechanism based on regional and continental simulations, *Atmospheric Environment*, 46, 417–429, <https://doi.org/10.1016/j.atmosenv.2011.09.030>, 2012.

NIST Chemistry WebBook: <https://webbook.nist.gov/chemistry/>, last access: 29 January 2024.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., and Bolton, E. E.: PubChem in 2021: new data content and improved web interfaces, *Nucleic Acids Research*, 49, D1388–D1395, <https://doi.org/10.1093/nar/gkaa971>, 2021.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., and Bolton, E. E.: PubChem 2023 update, *Nucleic Acids Research*, 51, D1373–D1380, <https://doi.org/10.1093/nar/gkac956>, 2023.

Koss, A. R., Sekimoto, K., Gilman, J. B., Selimovic, V., Coggon, M. M., Zarzana, K. J., Yuan, B., Lerner, B. M., Brown, S. S., Jimenez, J. L., Krechmer, J., Roberts, J. M., Warneke, C., Yokelson, R. J., and de Gouw, J.: Non-methane organic gas emissions from biomass burning: identification, quantification, and emission factors from PTR-ToF during the FIREX 2016 laboratory experiment, *Atmospheric Chemistry and Physics*, 18, 3299–3319, <https://doi.org/10.5194/acp-18-3299-2018>, 2018.

Li, M., Zhang, Q., Streets, D. G., He, K. B., Cheng, Y. F., Emmons, L. K., Huo, H., Kang, S. C., Lu, Z., Shao, M., Su, H., Yu, X., and Zhang, Y.: Mapping Asian anthropogenic emissions of non-

methane volatile organic compounds to multiple chemical mechanisms, *Atmospheric Chemistry and Physics*, 14, 5617–5638, <https://doi.org/10.5194/acp-14-5617-2014>, 2014.

May, A. A., McMeeking, G. R., Lee, T., Taylor, J. W., Craven, J. S., Burling, I., Sullivan, A. P., Akagi, S., Collett, J. L., Flynn, M., Coe, H., Urbanski, S. P., Seinfeld, J. H., Yokelson, R. J., and Kreidenweis, S. M.: Aerosol emissions from prescribed fires in the United States: A synthesis of laboratory and aircraft measurements, *Journal of Geophysical Research-Atmospheres*, 119, 11826–11849, <https://doi.org/10.1002/2014jd021848>, 2014.

Nannoolal, Y., Rarey, J., and Ramjugernath, D.: Estimation of pure component properties: Part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contributions and group interactions, *Fluid Phase Equilibria*, 269, 117–133, <https://doi.org/10.1016/j.fluid.2008.04.020>, 2008.

Pence, H. E. and Williams, A.: ChemSpider: An Online Chemical Information Resource, *J. Chem. Educ.*, 87, 1123–1124, <https://doi.org/10.1021/ed100697w>, 2010.

Radke, L. F., Hegg, D. A., Hobbs, P. V., Nance, J. D., Lyons, J. H., Laursen, K. K., Weiss, R. E., Riggan, P. J., and Ward, D. E.: Particulate and trace gas emissions from large biomass fires in North America, in: *Global biomass burning: atmospheric, climatic, and biospheric implications*, MIT Press, 209–224, 1991.

Selimovic, V., Yokelson, R. J., Warneke, C., Roberts, J. M., de Gouw, J., Reardon, J., and Griffith, D. W. T.: Aerosol optical properties and trace gas emissions by PAX and OP-FTIR for laboratory-simulated western US wildfires during FIREX, *Atmospheric Chemistry and Physics*, 18, 2929–2948, <https://doi.org/10.5194/acp-18-2929-2018>, 2018.

Stockwell, C. E., Veres, P. R., Williams, J., and Yokelson, R. J.: Characterization of biomass burning emissions from cooking fires, peat, crop residue, and other fuels with high-resolution proton-transfer-reaction time-of-flight mass spectrometry, *Atmospheric Chemistry and Physics*, 15, 845–865, <https://doi.org/10.5194/acp-15-845-2015>, 2015.

Topping, D., Barley, M., Bane, M. K., Higham, N., Aumont, B., Dingle, N., and McFiggans, G.: UManSysProp v1.0: an online and open-source facility for molecular property prediction and atmospheric aerosol calculations, *Geoscientific Model Development*, 9, 899–914, <https://doi.org/10.5194/gmd-9-899-2016>, 2016.

US EPA: Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11 or insert version used]. United States Environmental Protection Agency, Washington, DC, USA. 2023.

Yokelson, R. J., Susott, R., Ward, D. E., Reardon, J., and Griffith, D. W. T.: Emissions from smoldering combustion of biomass measured by open-path Fourier transform infrared spectroscopy, *Journal of Geophysical Research: Atmospheres*, 102, 18865–18877, <https://doi.org/10.1029/97JD00852>, 1997.

Yokelson, R. J., Christian, T. J., Karl, T. G., and Guenther, A.: The tropical forest and fire emissions experiment: laboratory fire measurements and synthesis of campaign data, *Atmos. Chem. Phys.*, 8, 3509–3527, <https://doi.org/10.5194/acp-8-3509-2008>, 2008.

Yokelson, R. J., Burling, I. R., Gilman, J. B., Warneke, C., Stockwell, C. E., de Gouw, J., Akagi, S. K., Urbanski, S. P., Veres, P., Roberts, J. M., Kuster, W. C., Reardon, J., Griffith, D. W. T., Johnson, T. J., Hosseini, S., Miller, J. W., Cocker, D. R., Jung, H., and Weise, D. R.: Coupling field and laboratory measurements to estimate the emission factors of identified and unidentified trace gases for prescribed fires, *Atmospheric Chemistry and Physics*, 13, 89–116, <https://doi.org/10.5194/acp-13-89-2013>, 2013.