Response from authors

Re: Review of Harmonized Emissions Component (HEMCO) 3.0 as a versatile emissions
component for atmospheric models: application in the GEOS-Chem, NASA GEOS, WRF-
GC, CESM2, NOAA GEFS-Aerosol, and NOAA UFS models

August 11, 2021

We thank the two Reviewers for their helpful comments. In response, we have carefully
revised the manuscript to clarify the current capabilities of HEMCO 3.0. We respond to each specific
comment in detail below. The referee comments are shown in red italics. Our replies are shown in
black and modified text is shown in blue. The annotated line and page numbers refer to the revised
copy of the manuscript.

A copy of the manuscript's complete changes is attached at the end of the response.

Reviewer #1

*I have read the paper "Harmonized Emissions Component (HEMCO) 3.0 as a versatile emissions
component for atmospheric models: application in the GEOS-Chem, NASA GEOS, WRF-GC,
CESM2, NOAA GEFS- Aerosol, and NOAA UFS models" by Lin et al. The paper we well written
and described the HEMCO component well. HEMCO 3.0 looks to be much more flexible and
useful for a wider range of models that the previous version.*
A few specific comments.
One point to clarify, can scaling be spatially-, temporally, or sector-explicit?
*(for example, sector-specific, or regionally-specific NOx -> NO2/NO2 scaling. Can this change
over time?) This seems to be implied latter in the text, but should be mentioned more explicitly.*

Thank you for your comment. Yes, scaling can be spatially, temporally, and sector-explicit. We have
revised the text to clarify this:

P6, Lines 168-171:

Scaling factors can be specified for each individual entry in the HEMCO configuration file, allowing for different scaling factors to be applied for different inventories, sectors, and species. HEMCO accepts scaling factors as constant numbers, temporally-explicit (diurnal, day-of-week, seasonal, interannual) numbers, or as a gridded netCDF data file.

There should be a software implementation section in the paper that gives an overview of how the software is implemented and structured. It appears the language is fortran90, but this is not mentioned anywhere in the text. It would be useful to provide a few paragraphs that give an overview of how the code is structured at the "nuts and bolts" level. How do the HEMCO components communicate with each other? Are there common bocks of data that are key to HEMCO (or have more modern software encapsulation and abstraction techniques been used?)

Thank you for your suggestion. We have added descriptions of the HEMCO 3.0 code in the manuscript.

P7, Lines 178-184:

The HEMCO Core calculates emissions with summations, masks, and scaling factors specified in the HEMCO configuration file. *It includes Fortran modules that define the HEMCO state, HEMCO data types (e.g., configuration options, date and time, chemical species and their physical properties, file containers storing input data, and data containers storing data processed by HEMCO), and the driver routine that computes emissions and stores them in data containers. All data types are contained in a variable called the HEMCO state (HcoState) and passed as an argument to all HEMCO subroutines in the code. This allows multiple instances of HEMCO to operate simultaneously, as multiple copies of HEMCO state can co-exist independently. The HEMCO Core also includes an error handling and logging component.*

P10, Lines 216-230:

At the beginning of the run, the Model Interface Layer provides the model species list to the HEMCO Core along with any physical properties needed for computation of state-dependent emissions (for example, Henry's law constants for ocean fluxes). It also provides information to the HEMCO Core on the model environment, such as the model clock and time step size. *This information is stored in the HEMCO state by the HEMCO Core.*

At every HEMCO time step, when HEMCO is called by the model, the HEMCO Core performs the requested calculations, loading the latest available input data *into the HEMCO state's file containers* using the Data Input Layer as necessary.

While the paper gives a useful overview of the way HEMCO is used, a reader comes away from the paper with no idea of how to actually use HEMCO? For example, is it a matter connecting to just a few key interface routines?

Thank you for your suggestion. We have added further detail regarding how to implement HEMCO 3.0 in a new model in the section regarding the Data Input Layer and the Model Interface Layer, which are used to "couple" HEMCO 3.0 to a new model.

P11, Lines 266-275:

Implementation of HEMCO in a new model can be prototyped by modifying the default Model Interface Layer to use the model's data structures. *The Model Interface Layer includes at least three subroutines: initialization, run, and finalization. These subroutines need to be called by the model and provided with information about the model environment. For initialization, the model species list and their physical properties, the HEMCO grid information, and the location of the configuration file need to be provided. For run, information about the current model time and model state variables for HEMCO extensions needs to be provided, and the computed emissions and data need to be retrieved from the HEMCO Core to be passed back to the model. If the HEMCO grid is different than the model*

grid (Sect. 2.4), then the Model Interface Layer also needs to implement re-gridding capabilities. An

example is HEMCO 3.0's implementation within the CESM model, described in Sect. 3.5.

As part of such a section, it would be very useful to provide some details on how a user would add

an extension. This is a potentially very powerful feature of the system, so more details on that aspect

would be useful. Give an indication of how much work someone would have to do to implement a

new extension. It would also be useful to point to some examples in the current code base to

extensions implemented in the current system.

Thank you for your suggestion. In response, we have added further information on how existing

extensions were implemented, how new extensions can be added in a similar manner, and a table

including the current HEMCO extensions available in HEMCO 3.0.

P9, Lines 200-211:

Emissions dependent on model state such as dust or lightning can be computed on-line by using

algorithms called HEMCO extensions supplied with the HEMCO Core. For example, the current

HEMCO Core includes as default the DEAD dust emission extension implementing the algorithm

from Zender et al. (2003) but users may select other available extension options (such as the Ginoux

et al. (2001) algorithm) or they can add a new algorithm as an extension. Alternatively, users may pre-

compute these emissions based on off-line input data and disable the HEMCO extension. Both

approaches are routinely used in GEOS-Chem (Weng et al., 2020). *Table 2 lists available HEMCO*

extensions in HEMCO 3.0. Users may add a new algorithm as an extension by creating a new

extension file within the Extensions directory in HEMCO. HEMCO extensions include subroutines for

initialization, run, and finalization. At every time step, the "Run" subroutine receives HEMCO state

and model state information, and returns the computed emissions array to the HEMCO Core, which

can then be added to the other emissions data. New state-dependent emission algorithms can be

modified to fit this structure by encapsulating the bulk of the code into the "Run" routine and

adjusting the variable names so that model state can be read through HEMCO. Most HEMCO

extensions were developed in this way.

Table 2. Emission extensions available in HEMCO 3.0 as built-in algorithms.

Species [a]	Extension name and reference
Oceanic DMS, Acetone, Acetaldehyde, Methyl nitrate, Ethyl nitrate, Methanol	SeaFlux
Ship plume NO_x, HNO_3, O_3	ParaNOx (Vinken et al., 2011)
Lightning NO_x	LightNOx (Murray et al., 2012; Ott et al., 2010)
Soil and fertilizer NO_x	SoilNOx (Hudman et al., 2012)
Mineral dust aerosols	DEAD (Zender et al., 2003)
	Ginoux (Ginoux et al., 2001)
Sea salt aerosols	SeaSalt (Chin et al., 2001; Gong, 2003; Jaeglé et al., 2011)
Biogenic VOCs	MEGAN (Guenther et al., 2012)
Biomass burning	GFED (Akagi et al., 2011; Andreae et al., 2001; Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010)
	FINN (Wiedinmyer et al., 2011)
Volcanic SO_2	Volcano (Carn et al., 2015; Ge et al., 2016)
Inorganic iodine emissions: HOI, I_2	Inorg_Iodine (Carpenter et al., 2013; MacDonald et al., 2014)

[a] DMS = dimethyl sulfide;

Reviewer #2

This manuscript describes the HEMCOv3 system, a flexible tool for processing and adapting emission inventories for air quality modelling purposes. The strength of HEMCO is in its ability to combine multiple inventories in a very flexible and transparent way, as well as in the fact that it can be used on-line with multiple atmospheric chemistry transport models. The paper is well structured and a good contribution to GMD. Nevertheless, it is sometimes lacking details. Therefore, the paper should be revised according to the following comments before being published.

The new version of HEMCO is capable of calculating emissions at any model resolution including multiscale and unstructured grids. This is highlighted by the authors as an improved capability of the system compared to its previous version. However, no description is provided in terms of: 1) how HEMCO creates multiscale and unstructured grids, 2) how the regridding of emissions is performed for these cases (i.e., how the mass conservation is ensured) and 3) how the definition of these types of grids needs to be provided by the user. Besides these points, authors should also clarify if the new version of HEMCO can perform a remapping to working grids using projections different than regular lat-lon (e.g., lambert conformal conic, polar stereographic) as these types of projections are widely used in the modelling community.

Thank you for your suggestions. The capability of HEMCO 3.0 to support other grids and projections is dependent on the implementations of the Data Input Layer and the Model Interface Layer, which provide re-gridding capabilities between the input data, the HEMCO grid (if different than the model grid), and the model grid. The HEMCO Core operates on grid boxes and is independent of the grid being used. HEMCO out-of-the-box includes the default Data Input Layer and Model Interface Layers, which support rectilinear latitude-longitude grids only. Other implementations of HEMCO 3.0 expand on this capability. For example, the GEOS-Chem High Performance (GCHP) implementation supports the cubed-sphere grid as the HEMCO grid through the *ExtData* component in MAPL, and the CESM2 implementation supports an arbitrary model grid described by an ESMF mesh file while

keeping a rectilinear latitude-longitude HEMCO grid. Future implementations may opt to use a mix of both approaches. We have updated the manuscript to clarify this.

P11, Lines 266-275:

Implementation of HEMCO in a new model can be prototyped by modifying the default Model Interface Layer to use the model's data structures. *The Model Interface Layer includes at least three subroutines: initialization, run, and finalization. These subroutines need to be called by the model and provided with information about the model environment. For initialization, the model species list and their physical properties, the HEMCO grid information, and the location of the configuration file need to be provided. For run, information about the current model time and model state variables for HEMCO extensions needs to be provided, and the computed emissions and data need to be retrieved from the HEMCO Core to be passed back to the model. If the HEMCO grid is different than the model grid (Sect. 2.4), then the Model Interface Layer also needs to implement re-gridding capabilities. An example is HEMCO 3.0's implementation within the CESM model, described in Sect. 3.5.*

P13, Lines 320-329:

While any unstructured grid may be used as the HEMCO grid, it may be desirable to use a rectilinear latitude-longitude grid for prototyping HEMCO in new models. This is because the default Data Input Layer provided with HEMCO only supports rectilinear latitude-longitude grids, and most input data available in the HEMCO database library are also on rectilinear latitude-longitude grids. By choosing such a grid, the default Data Input Layer can be readily used for quick prototyping of a new HEMCO implementation, which may then be improved upon if another HEMCO grid is more desirable. *In cases where other grids are used as the HEMCO grid or the model grid, conservative re-gridding needs to be implemented by the Data Input Layer or the Model Interface Layer. Examples of these scenarios are described in Sect. 3.2, where the HEMCO grid is a cubed-sphere grid and ExtData from MAPL with re-gridding capability is implemented as a Data Input Layer, and Sect. 3.5, where the model grid is an arbitrary grid described by an ESMF mesh file provided by the model to HEMCO, and ESMF on-line re-gridding is implemented in the Model Interface Layer.*

P19-20, Lines 462-469:

Figure 8 shows the implementation of HEMCO 3.0 in CESM2. HEMCO serves emissions to CAM-chem, GEOS-Chem, and potentially to any representation of atmospheric chemistry in CAM. The Model Interface Layer includes routines to export data processed by HEMCO to CAM's physics buffer, a temporary storage space for model components to share data at runtime. *As CAM supports a variety of grids, during initialization, the HEMCO-CESM Model Interface Layer reads the ESMF mesh file that describes the grid used by CAM, and uses ESMF on-line re-gridding (https://earthsystemmodeling.org/regrid/) to re-grid data between HEMCO and CAM. By using ESMF re-gridding capabilities, HEMCO is capable of operating independently of the grid used by the model, as long as the model grid description is provided to HEMCO using an ESMF mesh file.*

A general description of the "HEMCO extensions" should be provided (i.e., a brief description of how the built-in algorithms estimate emissions from vegetation, dust, lightning and ocean, including parameters and estimation methodologies used. Specific references to these extensions should be also included if available.

Thank you for your suggestion. We have added a paragraph describing the features of HEMCO extensions, how to implement them, and have included a table of the current extensions available in HEMCO 3.0.

P9, Lines 200-211:

Emissions dependent on model state such as dust or lightning can be computed on-line by using algorithms called HEMCO extensions supplied with the HEMCO Core. For example, the current HEMCO Core includes as default the DEAD dust emission extension implementing the algorithm from Zender et al. (2003) but users may select other available extension options (such as the Ginoux et al. (2001) algorithm) or they can add a new algorithm as an extension. Alternatively, users may pre-compute these emissions based on off-line input data and disable the HEMCO extension. Both

approaches are routinely used in GEOS-Chem (Weng et al., 2020). Table 2 lists available HEMCO

extensions in HEMCO 3.0. Users may add a new algorithm as an extension by creating a new

extension file within the Extensions directory in HEMCO. HEMCO extensions include subroutines for

initialization, run, and finalization. At every time step, the "Run" subroutine receives HEMCO state

and model state information, and returns the computed emissions array to the HEMCO Core, which

can then be added to the other emissions data. New state-dependent emission algorithms can be

modified to fit this structure by encapsulating the bulk of the code into the "Run" routine and

adjusting the variable names so that model state can be read through HEMCO. Most HEMCO

extensions were developed in this way.

Table 2. Emission extensions available in HEMCO 3.0 as built-in algorithms.

Species [a]	Extension name and reference
Oceanic DMS, Acetone, Acetaldehyde, Methyl nitrate, Ethyl nitrate, Methanol	SeaFlux
Ship plume NO_x, HNO_3, O_3	ParaNO$_x$ (Vinken et al., 2011)
Lightning NO_x	LightNO$_x$ (Murray et al., 2012; Ott et al., 2010)
Soil and fertilizer NO_x	SoilNO$_x$ (Hudman et al., 2012)
Mineral dust aerosols	DEAD (Zender et al., 2003) Ginoux (Ginoux et al., 2001)
Sea salt aerosols	SeaSalt (Chin et al., 2001; Gong, 2003; Jaeglé et al., 2011)
Biogenic VOCs	MEGAN (Guenther et al., 2012)
Biomass burning	GFED (Akagi et al., 2011; Andreae et al., 2001; Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010) FINN (Wiedinmyer et al., 2011)
Volcanic SO_2	Volcano (Carn et al., 2015; Ge et al., 2016)
Inorganic iodine emissions: HOI, I_2	Inorg_Iodine (Carpenter et al., 2013; MacDonald et al., 2014)

[a] DMS = dimethyl sulfide;

I think it would be useful for the reader to know which are the emission inventories that are currently considered in the database library of HEMCOv3. A summary table or link to a page where this information is specified would work.

Thank you for your suggestion. We have added Table 1, which includes the current emission inventories available in the HEMCO 3.0 library.

Page 8, Lines 189-193:

Table 1 lists the emission inventories currently in the HEMCO 3.0 default database library. Users can select from that list and can easily add new inventories. The hierarchy of emission inventories is specified in the configuration file. Masking is done by superseding lower hierarchy inventories with higher hierarchy inventories so that default inventories may be overwritten by different inventories available only for a particular region, period, or category. For example, an inventory specific for China in 2018 such as MEIC may overwrite a global default inventory.

Table 1. Emission inventories currently in the default HEMCO 3.0 library.

Coverage	Species[a]	Resolution (lat × lon)	Reference
Global	CO, NO_x, NH_3, SO_2, VOCs, BC, OC, CO_2	0.5° × 0.5°	CEDS (Hoesly et al., 2018; McDuffie et al., 2020)
Global	CO, NO_x, NH_3, SO_2, BC, OC	0.1° × 0.1°	EDGAR v4.3.2 (Crippa et al., 2018)
Global	CO, NO, NH_3, SO_2	0.1° × 0.1°	EDGAR-HTAP v2 (Janssens-Maenhout et al., 2012)
Global	Natural Sources NH_3	2° × 2.5°	GEIA (Bouwman et al., 1997)
Global	Arctic Seabird NH_3	0.25° × 0.25°	Croft et al. (2016)
Global	Anthropogenic C_2H_5OH	2° × 2.5°	Millet et al. (2010), Olivier et al. (2003), Granier et al. (2005)
Global	C_2H_6	4° × 5°	Tzompa-Sosa et al. (2017)
Global	C_3H_8	2° × 2.5°	Xiao et al. (2008)
Global	$CHBr_3$, CH_2Br_2	2° × 2.5°	Liang et al. (2010)

Global	CH_3I, CH_2I_2, CH_2ICl, CH_2IBr	$1° \times 1°$	Ordóñez et al. (2012), Sherwen et al. (2016)
Global	Aircraft CO, NO_x, SO_2, VOCs, BC, OC	$1° \times 1°$	AEIC (Stettler et al., 2011)
Global	CH_3CHO, C_2H_5OH	$2° \times 2.5°$	Millet et al. (2010)
Global	Anthropogenic $PM_{2.5}$ Dust	$2° \times 2.5°$	AFCID (Philip et al., 2017)
Global	Ship SO_2, CO, NO_x	$1° \times 1°$	ICOADS (Wang et al., 2008)
Global	Ship SO_2	$1° \times 1°$	ARCTAS (Eyring et al., 2005)
Global	Ship SO_2	$1° \times 1°$	Corbett et al. (1999)
Global	Future RCP3PD, RCP4.5, RCP6.0, RCP8.5 CO, NO_x, NH_3, SO_2, VOCs, BC, OC	$0.5° \times 0.5°$	Holmes et al. (2013)
Global[*]	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC	$0.25° \times 0.25°$	GFED v4.1 (Akagi et al., 2011; Andreae et al., 2001; Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010)
Global	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC, CO_2	$0.1° \times 0.1°$	GFAS (Di Giuseppe et al., 2018; Rémy et al., 2017; Andela et al., 2013; Kaiser et al., 2012, Xu et al., 2010; Heil et al., 2010; Di Giuseppe et al., 2016)
Global	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC, CO_2	$0.25° \times 0.25°$	QFED v2.5r1 (Darmenov et al., 2013)
Global	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC	$0.25° \times 0.25°$	BB4CMIP (van Merle et al., 2017)
US	CO, NO_x, NH_3, SO_2, VOCs, BC, OC	$0.1° \times 0.1°$	EPA (https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-data, last retrieved 23 Jul 2021)
Canada	CO, NO_x, NH_3, SO_2, BC, OC	$0.1° \times 0.1°$	APEI (https://pollution-waste.canada.ca/air-emission-inventory/, last retrieved 23 Jul 2021)
Africa	CO, NO_x, NH_3, SO_2, VOCs, BC, OC	$0.1° \times 0.1°$	DICE-Africa (Marais et al., 2016)
China	CO, NO_x, NH_3, SO_2, VOCs, BC, OC	$0.5° \times 0.66°$	MEIC (http://meicmodel.org/, last retrieved 23 Jul 2021; Li et al., 2017; Zheng et al., 2018)

[a] VOCs = volatile organic compounds; OC = organic carbon aerosol; BC = black carbon aerosol; * implemented as extension to distribute dry matter input data into model species.

Is there any pre-processing associated to the original emission inventories so that they can be ingested by HEMCO (e.g., renaming of pollutant names to have them homogenized across all inventories)?

Thank you for your question. The input data provided to HEMCO needs to be processed into COARDS-compliant netCDF format (http://wiki.seas.harvard.edu/geos-chem/index.php/The_COARDS_netCDF_conventions_for_earth_science_data).
The renaming of pollutant names is not necessary, as the HEMCO configuration file specifies the mapping between the netCDF variable name and the model chemical species name. This allows different models or chemical mechanisms to share the same HEMCO input files without further processing needed for each individual model. It may also be desirable, for performance reasons, to split the original inventory files by year or month, so they can be read faster during the model run. HEMCO will advance through the files automatically as the model time moves forward. We have revised the manuscript to include further detail on adding new inventories.

Page 6, Lines 143-156:

The HEMCO configuration file (example in Fig. 2) controls the operation of all HEMCO layers, fully describing the relationship between the input data read by the Data Input Layer, the processing by the HEMCO Core, and the data passed to the model by the Model Interface Layer. It is organized as individual entries for data, scaling factors, and masks. Each entry is numbered or named and includes information about the source of data (usually a netCDF file name but may be a number or mathematical expression in simple cases). *For netCDF data files, each entry specifies the netCDF variable name to be read (allowing the mapping from netCDF input species to model species), the temporal range, refresh frequency, cycling option (whether to continuously cycle the data or require an exact date match), and spatial dimension (2-D or 3-D data, with the option to specify a custom vertical distribution for 2-D data).* Also included is the model species name, the scaling factors to be applied, and the hierarchy (priority order used for masking). If a data entry does not include a species name, the entry is treated as generic data and is read into HEMCO, scaled, masked, and made

available to the model upon request. Entries may be organized in the form of "collections" enabled or disabled in bulk using switches. *HEMCO comes with a default database library of emission inventories and environmental data sets that is updated with every new GEOS-Chem version release, but users can readily add their own through processing the data into COARDS-compliant netCDF format and providing the corresponding configuration file entries. A detailed HEMCO User Guide is available on the GEOS-Chem Wiki (http://wiki.seas.harvard.edu/geos-chem/index.php/The_HEMCO_User%27s_Guide, retrieved 7 January 2021).*

Page 8-9, Lines 189-199:

Table 1 lists the emission inventories currently in the HEMCO 3.0 default database library. Users can select from that list and can easily add new inventories. The hierarchy of emission inventories is specified in the configuration file. Masking is done by superseding lower hierarchy inventories with higher hierarchy inventories so that default inventories may be overwritten by different inventories available only for a particular region, period, or category. For example, an inventory specific for China in 2018 such as MEIC may overwrite a global default inventory. A simulation for later years may retain the Chinese inventory for 2018, scale it up or down, or default to the global inventory, as specified in the configuration file. *Additional emission inventories can be added to the HEMCO library in COARDS-compliant netCDF format. The HEMCO configuration file allows HEMCO to remap variable names in the inventory source file to the species name in the model, specify the region that the inventory is used for, and its precedence over the existing entries in the HEMCO library, without necessitating pre-processing of the inventory source files.*

The vertical preprocessing of emissions is not required anymore in HEMCOv3, as it performs an on-line vertical regridding of the original emissions. It is however not clear to me how the user can assign sector-dependent vertical profiles to input emission inventories. This is relevant for those sectors in which emissions are originally reported in a single layer (e.g., emissions from energy sector in CEDS). Perhaps an example could be added in Figure 2.

Thank you for your suggestion. We have clarified the ability to apply scaling factors and specify custom vertical distributions in the manuscript, and have included an example in Fig. 2. Since each entry in the HEMCO configuration file is species-, sector-, and inventory-specific, it is possible to apply a custom scaling factor and vertical profile to a single species, sector, or inventory. This capability is used for distributing emissions vertically by sector in GEOS-Chem, CESM2-GC, and CAM-chem.

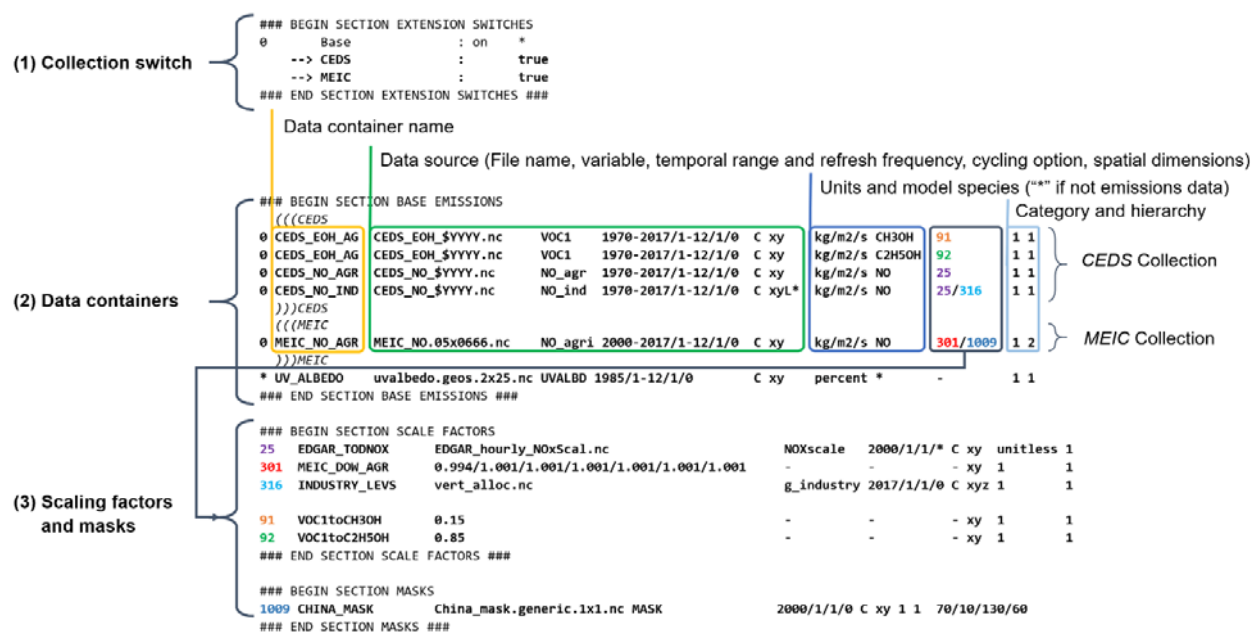



Figure 2. Sample HEMCO configuration file. The HEMCO configuration file is organized in three sections: (1) switches for "collections" of data containers, (2) data containers to be used in the model simulation, optionally organized into "collections", and (3) scaling and masking rules to be used. Entries are organized in a similar format, including a number and/or name, the data source (netCDF file and variable name, numbers, or mathematical expressions), the temporal range and spatial dimensions, and their category and hierarchy (in the same category, data entries with higher hierarchy take precedence). For data containers, scaling factors and masks are applied by referencing the numbered scaling factor and mask entries (colored text).

Page 6, Lines 143-156:

The HEMCO configuration file (example in Fig. 2) controls the operation of all HEMCO layers, fully describing the relationship between the input data read by the Data Input Layer, the processing by the HEMCO Core, and the data passed to the model by the Model Interface Layer. It is organized as individual entries for data, scaling factors, and masks. Each entry is numbered or named and includes information about the source of data (usually a netCDF file name but may be a number or

mathematical expression in simple cases). *For netCDF data files, each entry specifies the netCDF variable name to be read (allowing the mapping from netCDF input species to model species), the temporal range, refresh frequency, cycling option (whether to continuously cycle the data or require an exact date match), and spatial dimension (2-D or 3-D data, with the option to specify a custom vertical distribution for 2-D data).* Also included is the model species name, the scaling factors to be applied, and the hierarchy (priority order used for masking). If a data entry does not include a species name, the entry is treated as generic data and is read into HEMCO, scaled, masked, and made available to the model upon request. Entries may be organized in the form of "collections" enabled or disabled in bulk using switches. *HEMCO comes with a default database library of emission inventories and environmental data sets that is updated with every new GEOS-Chem version release, but users can readily add their own through processing the data into COARDS-compliant netCDF format and providing the corresponding configuration file entries. A detailed HEMCO User Guide is available on the GEOS-Chem Wiki (http://wiki.seas.harvard.edu/geos-chem/index.php/The_HEMCO_User%27s_Guide, retrieved 7 January 2021).*

P6, Lines 168-171:

Scaling factors can be specified for each individual entry in the HEMCO configuration file, allowing for different scaling factors to be applied for different inventories, sectors, and species. HEMCO accepts scaling factors as constant numbers, temporally-explicit (diurnal, day-of-week, seasonal, interannual) numbers, or as a gridded netCDF data file.

Page 10, Lines 229-243:

HEMCO 3.0 computes vertically distributed (3-D) emissions in the same way as 2-D. HEMCO 2.0 assumed the 72-level or the 47-level GEOS grid when reading vertically distributed emissions. This required pre-processing of the emission inventory files from their original vertical grids to the supported GEOS grid. Such pre-processing is not required anymore in HEMCO 3.0, which reads vertical emission data on any sigma-pressure grid described in the input netCDF inventory file. The input data are then vertically re-gridded on-line to the model vertical grid by the Data Input Layer

using the MESSy NCREGRID package (Jöckel, 2006). This functionality is used in the WRF-GC and CESM2 models, which have user-configurable vertical grids. *HEMCO 3.0 is also capable of distributing 2-D input data on a 3-D grid, by modifying the spatial dimension field in the HEMCO configuration file. An example is shown in Fig. 2 where the 2-D NO emission field from the CEDS inventory in the industrial sector is copied vertically to all levels using the flag "xyL*". Emissions may then be distributed using 3-D scaling factors read from a netCDF file, such as the scaling factor 316 in Fig. 2. It is additionally possible to emit all 2-D data to a particular level (e.g., to emit to level 5, using "xyL5") or distributed among an altitude range (e.g., "xyL=1:500m" or "xyL=1:PBL"). Emission heights can additionally be read from a netCDF file. The spatial dimension parameter can be applied individually to each entry in the HEMCO configuration file, and thus can be applied to an individual inventory, sector, or species. Detailed documentation of this capability is available on the HEMCO User's Guide (http://wiki.seas.harvard.edu/geos-chem/index.php/HEMCO_examples#Applying_2D_emissions_vertically, last retrieved 08 Aug 2021).*

HEMCOv3 can operate on a higher resolution than the model working grid for masking and scaling purposes. This feature allows achieving greater accuracy at country borders between different inventory domains. Authors illustrate this function with an example in which teo emission inventories are masked at resolutions of 4x5 deg and 2x2.5 deg, the later allowing a better resolution for the US-Mexico border. This is clear to me but, should not be better to directly perform the masking and scaling operations on the original grids of the emission inventories before performing the regridding? Nowadays most of the available global and regional emission inventories are provided at very fine resolutions (i.e., 0.1x0.1 deg), which would allow having a good definition of country borders. Some inventories (e.g., EMEP) even provide the information of the emitting country per grid cell. In that case, using a country mask would not be the best solution. Is the information of emitting country used in HEMCO when given?
P10 L268: There is a limit to the resolution of the HEMCO grid. Could you specify this limit?

Thank you for your questions. We have revised the manuscript to clarify the limit and application of the HEMCO grid. It is necessary to balance the computational requirements needed to process data at a higher resolution and the desired accuracy. When multiple inventories are used, the inventories need to be re-gridded to a common HEMCO grid for processing, as the capability to perform masking by the hierarchy and category of different entries in the HEMCO configuration file has to run on a single grid. For a global model, it may be computationally infeasible to perform the masking and scaling at inventories' native resolution, but this is feasible when running a regional model or with HEMCO off-line.

Page 12-13, Lines 295-319:

Figure 3 illustrates the benefit of using a finer HEMCO grid at the boundaries between inventories. When the HEMCO grid is disabled, HEMCO runs at model resolution, and all input data, including masks, are re-gridded by the Data Input Layer to the model resolution before emissions are computed by the HEMCO Core. In the example of Fig. 3, where a national inventory for the US is to overwrite a global default inventory, this overwriting can be done only for grid cells that are fully in the US. Grid cells straddling the border must retain the global default in order to avoid under- or over-accounting, but this then loses information from the US inventory. *This is not a problem if the national inventory straddles the border and includes information on the fractional contributions from the neighboring country, but such is not the case here.* Using a finer, intermediate-resolution grid – the HEMCO grid – allows emissions at the model grid scale to more accurately blend the contributions from the two sides of the border in a single *model* grid cell. This also ensures greater consistency when using the same model simulations at different resolutions. As long as the HEMCO grid is kept at a single resolution, the calculated emissions will be consistent between simulations – no matter what model grid resolution is selected.

Another advantage of using a finer HEMCO grid is for emissions computed with extensions and dependent on both the model variables (provided on the model grid) and environmental data (provided on the HEMCO grid). If there is non-linear dependence of emissions on the environmental

data variables, then a finer HEMCO grid will produce more accurate emissions. This is the case for example in dust emission algorithms that use land type as a categorical variable.

There is a limit to the resolution of the HEMCO grid because of the need for HEMCO to store the different inventories in memory, *re-grid the data to the HEMCO grid*, and process the data at higher resolution, which may be computationally expensive. *In global simulations using GEOS-Chem Classic on a single machine at resolutions of 4^ox5^o or 2^ox2.5^o we have found that a HEMCO grid of 1^ox1.25^o is a practical limit. However, masking on the native inventory grid at much higher resolution may be desirable for regional modeling applications. One can circumvent the problem by pre-processing the emissions on their native grids using HEMCO in off-line mode. Another option is to use a regional rather than global model. For example, in nested GEOS-Chem Classic simulations at 0.25^ox0.3125^o resolution we find that HEMCO can easily handle a HEMCO grid of 0.1^ox0.1^o, typical of the native resolution of inventories.*

Figure 3: "Using the 4 x finer" – Should not be 2x finer?

Thank you for pointing this out. We have removed the "4x" as the coarse resolution ($4° \times 5°$) and fine resolution ($2° \times 2.5°$) provide sufficient information.

Harmonized Emissions Component (HEMCO) 3.0 as a versatile emissions component for atmospheric models: application in the GEOS-Chem, NASA GEOS, WRF-GC, CESM2, NOAA GEFS-Aerosol, and NOAA UFS models

5 Haipeng Lin[1], Daniel J. Jacob[1], Elizabeth W. Lundgren[1], Melissa P. Sulprizio[1], Christoph A. Keller[2,3], Thibaud M. Fritz[4], Sebastian D. Eastham[4], Louisa K. Emmons[5], Patrick C. Campbell[6,7], Barry Baker[7,8], Rick D. Saylor[9], Raffaele Montuoro[10,11]

[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[2]Universities Space Research Association, Columbia, MD, USA
10 [3]NASA Global Modeling and Assimilation Office, Greenbelt, MD, USA
[4]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA
[5]Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO, USA
[6]Center for Spatial Information Science and Systems, Cooperative Institute for Satellite Earth System Studies (CISESS), 15 George Mason University, Fairfax, VA, USA
[7]NOAA Air Resources Laboratory Affiliate, College Park, MD, USA
[8]Joint Center for Earth Systems Technology (JCET), University of Maryland Baltimore County, Baltimore, MD, USA
[9]NOAA Air Resources Laboratory, Atmospheric Turbulence and Diffusion Division, Oak Ridge, TN, USA
[10]Cooperative Institute for Research in Environmental Sciences, CU Boulder, Boulder, CO, USA
20 [11]NOAA Global Systems Laboratory, Boulder, CO, USA

Correspondence to: Haipeng Lin (hplin@seas.harvard.edu)

Abstract. Emissions are a central component of atmospheric chemistry models. The Harmonized Emissions Component (HEMCO) is a software component for computing emissions from a user-selected ensemble of emission inventories and 25 algorithms. ~~While available in standalone mode, HEMCO also provides a general on-line facility for models to compute emissions at runtime.~~ It allows users to re-grid, combine, overwrite, subset, and scale emissions from different inventories through a configuration file and with no change to the model source code. The configuration file also maps emissions to model species with appropriate units. HEMCO can operate in off-line standalone mode, but more importantly it provides an on-line facility for models to compute emissions at runtime. HEMCO complies with the Earth System Modeling Framework (ESMF) 30 for portability across models. We present here a new version HEMCO 3.0 that features an improved three-layer architecture to facilitate implementation into any atmospheric model, and improved capability for calculating emissions at any model resolution including multiscale and unstructured grids. The three-layer architecture of HEMCO 3.0 includes (1) a Data Input Layer that reads the configuration file and accesses the HEMCO library of emission inventories and other environmental data; (2) the HEMCO Core that computes emissions on the user-selected HEMCO grid; and (3) the Model Interface Layer that re-35 grids (if needed) and serves the data to the atmospheric model, and also serves model data to the HEMCO Core for computing

emissions dependent on model state (such as from dust, vegetation, etc.). The HEMCO Core is common to the implementation in all models, while the Data Input Layer and the Model Interface Layer are adaptable to the model environment. Default versions of the Data Input Layer and Model Interface Layer enable straightforward implementation of HEMCO in any simple model architecture, and options are available to disable features such as re-gridding that may be done by independent couplers in more complex architectures. The HEMCO library of emission inventories and algorithms is continuously enriched through user contributions, so that new inventories can be immediately shared across models. HEMCO can also serve as a general data broker for models to process input data not only for emissions but for any gridded environmental datasets. We describe existing implementations of HEMCO 3.0 in (1) the GEOS-Chem 'Classic' chemical transport model with shared-memory infrastructure, (2) the high-performance GEOS-Chem (GCHP) model with distributed-memory architecture, (3) the NASA GEOS Earth System Model (GEOS ESM), (4) the Weather Research and Forecasting model with GEOS-Chem (WRF-GC), (5) the Community Earth System Model Version 2 (CESM2), and (6) the NOAA Global Ensemble Forecast System – Aerosols (GEFS-Aerosols), and the planned implementation in the NOAA Unified Forecast System (UFS). Implementation of HEMCO in the CESM2 model contributes to the Multi-Scale Infrastructure for Chemistry and Aerosols (MUSICA) by providing a common emissions infrastructure to support different simulations of atmospheric chemistry across scales.

1 Introduction

Emissions are a crucial component in modeling of atmospheric chemistry. Models apply emission fluxes calculated from inputs including gridded inventory data, point source data, and environmental data. These data originate from an ensemble of sources with different spatiotemporal resolution and extent, and covering different chemical species. They may need to be re-gridded, combined, overlaid, scaled, or extended through computational algorithms to produce the model emissions. Here we present the Harmonized Emission Component (HEMCO) 3.0 as a versatile tool to ingest and process emission data in atmospheric models and share these data across models.

Emissions can be computed in atmospheric models either off-line or on-line. An off-line emissions module pre-computes emissions on the target model grid and archives them as time-varying files for input to the model. An on-line emissions module computes emissions at runtime within the model from a set of input files containing emission information and with rules for how this information is to be used. Off-line processing of emissions is used by many models, such as the PREP-CHEM-SRC pre-processor system (Freitas et al., 2011) used within the WRF-Chem model (Grell et al., 2005; Fast et al., 2006) and the Sparse Matrix Operator Kernel Emissions system (SMOKE, https://www.cmascenter.org/smoke/) used in models such asthe CMAQ (Byun and Schere, 2006) and CAMx (Ramboll Environment and Health, 2020).) models. The advantage of computing emissions at the pre-processing stage is the versatility in pre-processing tools and low computational requirements at runtime. However, the pre-processing of emissions is a cumbersome step and the resulting emission files may be prohibitively large

(Jähn et al., 2020). Any change to the emissions requires re-running the pre-processor code. Emissions dependent on environmental variables computed in the atmospheric model cannot be treated off-line, complicating the infrastructure.

70   The Harmonized Emissions Component (HEMCO), originally developed by Keller et al. (2014) and formerly called the Harvard-NASA Emissions Component, computes emissions customized to user needs through a configuration file and a database library. Atmospheric chemistry modelers can use HEMCO either off-line in a standalone mode to compute and archive emissions, or on-line to compute emissions at runtime and serve them to the model at each time step. HEMCO can select, modify, re-grid, combine, and supersede emission inventories and algorithms without changing the model source code

75   ~~of their model~~. Built-in algorithms called 'extensions' compute emissions dependent on environmental data and model state variables such as for vegetation, dust, lightning, and oceans. Subgrid processing of emissions to account for fast chemistry, as in ship plumes (Vinken et al., 2011), is also done in extensions. Selection of HEMCO extensions is done in the configuration file and the computations are done independently of the atmospheric model, allowing for immediate portability to other models.

80   HEMCO was originally developed for the GEOS-Chem atmospheric chemistry model (Bey et al., 2001; Eastham et al., 2018), where the current version HEMCO 2.0 has two different implementations. The 'Classic' version of GEOS-Chem with single-node shared-memory parallelization (OpenMP) and rectilinear latitude-longitude grids (Bey et al., 2001) uses HEMCO to its full extent including reading, re-gridding, and temporal interpolation of input files. In GEOS-Chem 'Classic', HEMCO is used not only for emissions but as a general data broker to read and process all model input data, including meteorological fields

85   and initial conditions. This has enabled in particular the FlexGrid algorithm to run GEOS-Chem on any custom nest selected at runtime (Li et al., 2021). This implementation of HEMCO is also used in the recently developed coupling of GEOS-Chem with the Weather Research and Forecasting (WRF) model (WRF-GC; Lin et al., 2020, Feng et al., 2021).

The "High-Performance" version of GEOS-Chem (GCHP; Eastham et al., 2018) ~~used~~uses a different implementation of

90   HEMCO 2.0. GCHP is designed for multi-node massively parallel computation using a distributed-memory parallelization (MPI) enabled by the NASA Model Analysis and Prediction Layer (MAPL, Suarez et al., 2007) which acts as the model's infrastructure and handles inter-node communication. MAPL is built upon the Earth System Modeling Framework (ESMF, Hill et al., 2004) and handles data read, re-gridding, and interpolation, so the corresponding HEMCO routines are disabled. This implementation of HEMCO is also presently used by the GOCART aerosol model operating within the MAPL-based

95   NASA Goddard Earth Observing System (GEOS) Earth Science model (Rienecker et al., 2008; Randles et al., 2017).

HEMCO 2.0 has several limitations that limit its portability to other models. First, the re-gridding capability is limited to latitude-longitude grids. Second, its implementation ~~of HEMCO~~ in distributed-memory environments uses MAPL-specific features. Third, it has a multiplicity of model access points that introduce unnecessary dependency on model code. Fourth, it

100   requires that emissions be computed on the model grid, which may introduce inaccuracies in masking of regional inventories

and in non-linear computations, and further necessitates duplicate copies of HEMCO to handle different resolutions in multi-scale model applications such as WRF-GC.

HEMCO 3.0 overcomes all these limitations of HEMCO 2.0. Construction of HEMCO 3.0 was motivated by interest from the Community Earth System Model Version 2 (CESM2; Pfister et al., 2020) and the NOAA Unified Forecast System (UFS; Campbell et al., 2020) in using HEMCO as an emissions component. This led us to develop a more modularized and powerful structure to increase accuracy and portability to different atmospheric models including with multi-scale and unstructured grids. HEMCO 3.0 preserves a shared common core for calculating emissions by selecting, adding, superseding (masking), and scaling emission inventories as specified by the user. Other parts of HEMCO are modularized to facilitate the incorporation of HEMCO into the specific software environment of the target model. A three-layer architecture is created to separate (1) input and re-gridding of data, (2) emission calculations using the HEMCO Core, and (3) coupling to the target model, including export of the computed emissions and import of model state variables for state-dependent emissions (extensions). With this new modularity and flexibility, HEMCO can be readily implemented in a wide range of model environments. Use of a common HEMCO Core facilitates the sharing of emission data and algorithms between models and the intercomparisons of model results.

2 HEMCO 3.0 description

2.1 General architecture

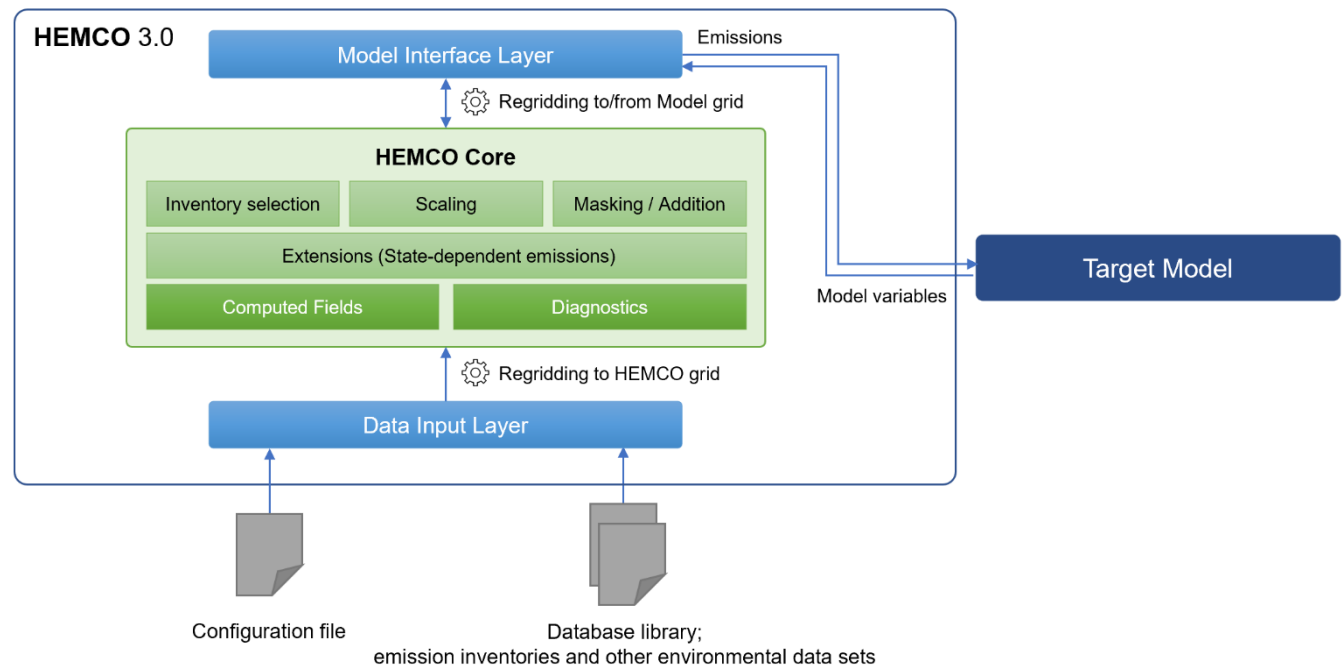



Figure 1. Three-layer architecture of HEMCO 3.0 including the Data Input Layer, the HEMCO Core, and the Model Interface Layer.

120

HEMCO 3.0 is modularized into a three-layer architecture as shown in Fig. 1, consisting of a Data Input Layer, the HEMCO Core, and the Model Interface ~~L ayer~~Layer. The Data Input Layer reads the configuration file and the database library of emission inventories and other environmental information, and re-grids the data to a user-defined HEMCO grid (finer than or identical to the model grid). The HEMCO Core assembles the emissions on the basis of instructions in the configuration file

125   including adding, scaling, and masking of individual inventories, and computing of emissions dependent on model state variables and environmental data (through algorithms referred to as HEMCO extensions). The Model Interface Layer communicates HEMCO output (including ~~computed fields such as~~ emission fluxes ~~and~~, diagnostics, and other computed data) to the target atmospheric model (hereafter referred to as "the model"), with re-gridding to the model grid if needed, and takes in and re-grids model variables to the HEMCO grid for use in HEMCO extensions. The Data Input Layer and the Model

130   Interface Layer have different implementations depending on the architecture of the model. The HEMCO Core, where emissions are computed, is the same in all cases. HEMCO operates on a horizontal "HEMCO grid" that may be any user-desired grid configuration (e.g., finer than the model grid, or the finest element of a multi-scale model grid), with the other layers handling the re-gridding to/from the HEMCO grid as necessary. 2-D (horizontal) emissions can be released at the surface or allocated vertically on the model grid as specified by the user through the HEMCO configuration file. 3-D emission

135   databases (such as for aircraft emissions) are re-gridded vertically to the model grid through the Data Input Layer.

Figure 2. Sample HEMCO configuration file. The HEMCO configuration file is organized in three sections: (1) switches for "collections" of data containers, (2) ~~the~~ data containers to be used in the model simulation, optionally organized into "collections", and (3) ~~the~~ scaling and masking rules to be used. Entries ~~in both sections~~ are organized in a similar format, including a number and/or name, the data source (netCDF file and variable name, numbers, or mathematical expressions), the temporal range ~~options~~ and spatial dimensions, and their category and hierarchy (in the same category, data entries with higher hierarchy take precedence). For data containers, scaling factors and masks are applied by referencing the numbered scaling factor and mask entries (colored text).

The HEMCO configuration file (example in Fig. 2) controls the operation of all HEMCO layers, fully describing the relationship between the input data read by the Data Input Layer, the processing by the HEMCO Core, and the data passed to the model by the Model Interface Layer. It is organized as individual entries for data, scaling factors, and masks. Each entry is numbered or named and includes information about the source of data (usually a netCDF file name but may be a number or mathematical expression in simple cases). For netCDF data files, each entry specifies the netCDF variable name to be read (allowing the mapping from netCDF input species to model species), the temporal range, refresh frequency, cycling option (whether to continuously cycle the data or require an exact date match~~)~~), and spatial dimension (2-D or 3-D data~~). This information is used by~~, with the ~~Data Input Layer. It also includes~~option to specify a custom vertical distribution for 2-D data). Also included is the model species name, the scaling factors to be applied, and the hierarchy (priority order used for masking). If a data entry does not include a species name, the entry is treated as generic data and is read into HEMCO, scaled, masked, and made available to the model upon request. ~~This information is used by the HEMCO Core for processing and by the Model Interface Layer for export to the model. For organization and easy control of inventories, entries~~Entries may be organized in the form of "collections~~", which can be~~" enabled or disabled in bulk using switches. HEMCO comes with a default database

160   library of emission inventories and environmental data sets ~~to which~~that is updated with every new GEOS-Chem version release, but users can readily add their own. ~~A detailed specification of~~ through processing the ~~HEMCO~~ data into COARDS-compliant netCDF format and providing the corresponding configuration file entries. A detailed HEMCO User Guide is available on the GEOS-Chem Wiki (http://wiki.seas.harvard.edu/geos-chem/index.php/The_HEMCO_User%27s_Guide, retrieved 7 January 2021).

165

    The HEMCO configuration file makes it possible for different models with different chemical species to share a single set of emissions input data (the "HEMCO database library"), without manually pre-processing the files for each mechanism. The variable name option in each entry allows for the mapping of the species name in the netCDF file to the model species name. One can also partition a class of species from the inventory into individual model species. For example, total alcohols in the

170   CEDS inventory (Hoesly et al., 2018) are to be emitted as 15% methanol and 85% ethanol in the CAM-chem model (Emmons et al., 2020). In that example, as illustrated in Fig. 2, HEMCO scaling factors are used to scale the same input variable by 15% and 85% into the CH3OH and C2H5OH model species. The scaling factor functionality in HEMCO can be used for temporal scaling (diurnal, day-of-week, seasonal, interannual) or to convert units from the emission inventory in the HEMCO database library to the target model. For example, emissions may be provided as kilograms of NO_x on an NO_2 mass basis in the inventory

175   file, but emitted as NO and NO_2 in the model. Scaling factors can be specified for each individual entry in the HEMCO configuration file, allowing for different scaling factors to be applied for different inventories, sectors, and species. HEMCO accepts scaling factors as constant numbers, temporally-explicit (diurnal, day-of-week, seasonal, interannual) numbers, or as a gridded netCDF data file.

180   The Data Input Layer processes each enabled entry in the HEMCO configuration file, reads the corresponding files from disk, and re-grids them to the HEMCO grid. The Data Input Layer then passes the data to the HEMCO Core in the form of data containers corresponding to each entry in the configuration file.

2.2 HEMCO Core

185   The HEMCO Core calculates emissions with summations, masks, and scaling factors specified in the HEMCO configuration file. ~~Masking is done by specifying an inventory hierarchy,~~ It includes Fortran modules that define the HEMCO state, HEMCO data types (e.g., configuration options, date and time, chemical species and their physical properties, file containers storing input data, and data containers storing data processed by HEMCO), and the driver routine that computes emissions and stores them in data containers. All data types are contained in a variable called the HEMCO state (HcoState) and passed as an

190   argument to all HEMCO subroutines in the code. This allows multiple instances of HEMCO to operate simultaneously, as multiple copies of HEMCO state can co-exist independently. The HEMCO Core also includes an error handling and logging component.

Table 1. Emission inventories currently in the default HEMCO 3.0 library.

Coverage	Species[a]	Resolution (lat × lon)	Reference
Global	Anthropogenic and ship CO, NO_x, NH_3, SO_2, VOCs, BC, OC, CO_2	0.5° × 0.5°	CEDS (Hoesly et al., 2018; McDuffie et al., 2020)
Global	Anthropogenic CO, NO_x, NH_3, SO_2, BC, OC	0.1° × 0.1°	EDGAR v4.3.2 (Crippa et al., 2018)
Global	Anthropogenic CO, NO, NH_3, SO_2	0.1° × 0.1°	EDGAR-HTAP v2 (Janssens-Maenhout et al., 2012)
Global	Natural sources NH_3	2° × 2.5°	GEIA (Bouwman et al., 1997)
Global	Arctic seabird NH_3	0.25° × 0.25°	Croft et al. (2016)
Global	Anthropogenic C_2H_5OH	2° × 2.5°	Millet et al. (2010), Olivier et al. (2003), Granier et al. (2005)
Global	Fossil fuel and biofuel C_2H_6	4° × 5°	Tzompa-Sosa et al. (2017)
Global	C_3H_8	2° × 2.5°	Xiao et al. (2008)
Global	$CHBr_3$, CH_2Br_2	2° × 2.5°	Liang et al. (2010)
Global	CH_3I, CH_2I_2, CH_2ICl, CH_2IBr	1° × 1°	Ordóñez et al. (2012), Sherwen et al. (2016)
Global	Aircraft CO, NO_x, SO_2, VOCs, BC, OC	1° × 1°	AEIC (Stettler et al., 2011)
Global	CH_3CHO, C_2H_5OH	2° × 2.5°	Millet et al. (2010)
Global	Anthropogenic $PM_{2.5}$ Dust	2° × 2.5°	AFCID (Philip et al., 2017)
Global	Ship SO_2, CO, NO_x	1° × 1°	ICOADS (Wang et al., 2008)
Global	Ship SO_2	1° × 1°	ARCTAS (Eyring et al., 2005)
Global	Ship SO_2	1° × 1°	Corbett et al. (1999)
Global	Future RCP3PD, RCP4.5, RCP6.0, RCP8.5 CO, NO_x, NH_3, SO_2, VOCs, BC, OC	0.5° × 0.5°	Holmes et al. (2013)
Global[*]	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC	0.25° × 0.25°	GFED v4.1 (Akagi et al., 2011; Andreae et al., 2001; Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010)

Global	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC, CO_2	0.1° × 0.1°	GFAS (Di Giuseppe et al., 2018; Rémy et al., 2017; Andela et al., 2013; Kaiser et al., 2012, Xu et al., 2010; Heil et al., 2010; Di Giuseppe et al., 2016)
Global	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC, CO_2	0.25° × 0.25°	QFED v2.5r1 (Darmenov et al., 2013)
Global	Fire CO, NO_x, NH_3, SO_2, VOCs, BC, OC	0.25° × 0.25°	BB4CMIP (van Merle et al., 2017)
US	CO, NO_x, NH_3, SO_2, VOCs, BC, OC	0.1° × 0.1°	EPA (https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-data, last retrieved 23 Jul 2021)
Canada	CO, NO_x, NH_3, SO_2, BC, OC	0.1° × 0.1°	APEI (https://pollution-waste.canada.ca/air-emission-inventory/, last retrieved 23 Jul 2021)
Africa	CO, NO_x, NH_3, SO_2, VOCs, BC, OC	0.1° × 0.1°	DICE-Africa (Marais et al., 2016)
China	CO, NO_x, NH_3, SO_2, VOCs, BC, OC	0.5° × 0.66°	MEIC (http://meicmodel.org/, last retrieved 23 Jul 2021; Li et al., 2017; Zheng et al., 2018)

195   [a] VOCs = volatile organic compounds; OC = organic carbon aerosol; BC = black carbon aerosol; * implemented as extension to distribute dry matter input data into model species.

Table 1 lists the emission inventories currently in the HEMCO 3.0 default database library. Users can select from that list and can easily add new inventories. The hierarchy of emission inventories is specified in the configuration file. Masking is done

200   by superseding lower hierarchy inventories with higher hierarchy inventories so that default inventories may be overwritten by different inventories available only for a particular region or, period., or category. For example, an inventory specific for China in 2018 such as MEIC may overwrite a global default inventory. A simulation for later years may retain the Chinese inventory for 2018, scale it up or down, or default to the global inventory, as specified in the configuration file. Additional emission inventories can be added to the HEMCO library in COARDS-compliant netCDF format. The HEMCO configuration

205   file allows HEMCO to remap variable names in the inventory source file to the species name in the model, specify the region that the inventory is used for, and its precedence over the existing entries in the HEMCO library, without necessitating pre-processing of the inventory source files.

Emissions dependent on model state such as dust or lightning ~~may~~can be computed on-line ~~in the HEMCO Core~~ by using algorithms called HEMCO extensions supplied with the HEMCO Core. For example, the current HEMCO Core includes as default the DEAD dust emission extension implementing the algorithm from Zender et al. (2003) but users may select other available extension options (such as the Ginoux et al. (2001) algorithm) or they can add a new algorithm as an extension. Alternatively, users may pre-compute ~~their~~these emissions based on off-line input data and disable the HEMCO extension. Both approaches are routinely used in GEOS-Chem (Weng et al., 2020). Table 2 lists available HEMCO extensions in HEMCO 3.0. Users may add a new algorithm as an extension by creating a new extension file within the Extensions directory in HEMCO. HEMCO extensions include subroutines for initialization, run, and finalization. At every time step, the "Run" subroutine receives HEMCO state and model state information, and returns the computed emissions array to the HEMCO Core, which can then be added to the other emissions data. New state-dependent emission algorithms can be modified to fit this structure by encapsulating the bulk of the code into the "Run" routine and adjusting the variable names so that model state can be read through HEMCO. Most HEMCO extensions were developed in this way.

Table 2. Emission extensions available in HEMCO 3.0 as built-in algorithms.

Species [a]	Extension name and reference
Oceanic DMS, Acetone, Acetaldehyde, Methyl nitrate, Ethyl nitrate, Methanol	SeaFlux
Ship plume NO_x, HNO_3, O_3	ParaNO$_x$ (Vinken et al., 2011)
Lightning NO_x	LightNO$_x$ (Murray et al., 2012; Ott et al., 2010)
Soil and fertilizer NO_x	SoilNO$_x$ (Hudman et al., 2012)
Mineral dust aerosols	DEAD (Zender et al., 2003)
	Ginoux (Ginoux et al., 2001)
Sea salt aerosols	SeaSalt (Chin et al., 2001; Gong, 2003; Jaeglé et al., 2011)
Biogenic VOCs	MEGAN (Guenther et al., 2012)
Biomass burning	GFED (Akagi et al., 2011; Andreae et al., 2001; Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010)
	FINN (Wiedinmyer et al., 2011)
Volcanic SO_2	Volcano (Carn et al., 2015; Ge et al., 2016)
Inorganic iodine emissions: HOI, I_2	Inorg_Iodine (Carpenter et al., 2013; MacDonald et al., 2014)

[a] DMS = dimethyl sulfide;

225 At the beginning of the run, the Model Interface Layer provides the model species list to the HEMCO Core along with any physical properties needed for computation of state-dependent emissions (for example, Henry's law constants for ocean fluxes). It also provides information to the HEMCO Core on the model environment, such as the model clock and time step size. This information is stored in the HEMCO state by the HEMCO Core.

230 At every HEMCO time step, when HEMCO is called by the model, the HEMCO Core performs the requested calculations, loading the latest available input data into the HEMCO state's file containers using the Data Input Layer as necessary. Emission fluxes are summed by species, and non-emissions data are stored individually by their data container name (UV albedo example in Fig. 2). The Model Interface Layer then exports the computed fields to the model, interpolating the data to the model grid if it is different than the HEMCO grid. The Model Interface Layer also passes updated model state information to the HEMCO

235 Core for use in extensions (for example, wind speeds to calculate dust emissions). That model state information is re-gridded to the HEMCO grid for purpose of the extension computations.

HEMCO 3.0 computes vertically distributed (3-D) emissions in the same way as 2-D. HEMCO 2.0 assumed the 72-level or the 47-level GEOS grid when reading vertically distributed emissions. This required pre-processing of the emission inventory

240 files from their original vertical grids to the supported GEOS grid. Such pre-processing is not required anymore in HEMCO 3.0, which reads vertical emission data on any sigma-pressure grid described in the input netCDF inventory file. The input data are then vertically re-gridded on-line to the model vertical grid by the Data Input Layer using the MESSy NCREGRID package (Jöckel, 2006). This functionality is used in the WRF-GC and CESM2 models, which have user-configurable vertical grids. HEMCO 3.0 is also capable of distributing 2-D input data on a 3-D grid, by modifying the spatial dimension field in the

245 HEMCO configuration file. An example is shown in Fig. 2 where the 2-D NO emission field from the CEDS inventory in the industrial sector is copied vertically to all levels using the flag "xyL*". Emissions may then be distributed using 3-D scaling factors read from a netCDF file, such as the scaling factor 316 in Fig. 2. It is additionally possible to emit all 2-D data to a particular level (e.g., to emit to level 5, using "xyL5") or distributed among an altitude range (e.g., "xyL=1:500m" or "xyL=1:PBL"). Emission heights can additionally be read from a netCDF file. The spatial dimension parameter can be applied

250 individually to each entry in the HEMCO configuration file, and thus can be applied to an individual inventory, sector, or species. Detailed documentation of this capability is available on the HEMCO User's Guide (http://wiki.seas.harvard.edu/geos-chem/index.php/HEMCO_examples#Applying_2D_emissions_vertically, last retrieved 08 Aug 2021).

HEMCO also includes as diagnostic capability a netCDF output component to archive selected emissions at specfied time

255 steps to disk. This may also include custom diagnostic quantities, such as lightning flash rate from the Lightning NO_x extension. The output component is used when HEMCO operates off-line in standalone mode, and also in GEOS-Chem Classic where HEMCO handles emission diagnostics. Other target models may have their own diagnostic packages on the model grid, in which case HEMCO diagnostics can be disabled.

2.3 Default Data Input Layer and Model Interface Layer

260 HEMCO 3.0 includes default out-of-the-box implementations for the Data Input Layer and the Model Interface Layer to enable simple implementation in new models without the need for specific information on model architecture. These default implementations are the ones used in the interface with GEOS-Chem 'Classic' and minimize dependencies on external libraries, facilitating application in a new model environment but with limited features. More advanced implementations are often desirable and will be presented in Sect. 3.

265

The HEMCO default Data Input Layer is a netCDF input component with rectilinear latitude-longitude grid re-gridding capabilities. It can be used out-of-the-box in HEMCO with no additional software dependencies. However, it does not support parallel input and as such may be inefficient in a massively parallel model environment. In models which use a non-rectilinear grid for computation or data, it would be necessary to ~~replace or~~ modify the default Data Input Layer if re-gridding is to be 270 performed.

The HEMCO default Model Interface Layer is a module with common utilities for the model to interface with HEMCO. It allows the model to retrieve emission fluxes from HEMCO and control HEMCO subroutines. As the Model Interface Layer is the point of access for the model to interact with HEMCO, implementation of HEMCO in a new model requires a new Model 275 Interface Layer to provide HEMCO with information on the model environment and updating model state information for HEMCO extensions.

Implementation of HEMCO in a new model can be prototyped by modifying the default Model Interface Layer to use the model's data structures. The Model Interface Layer includes at least three subroutines: initialization, run, and finalization. 280 These subroutines need to be called by the model and provided with information about the model environment. For initialization, the model species list and their physical properties, the HEMCO grid information, and the location of the configuration file need to be provided. For run, information about the current model time and model state variables for HEMCO extensions needs to be provided, and the computed emissions and data need to be retrieved from the HEMCO Core to be passed back to the model. If the HEMCO grid is different than the model grid (Sect. 2.4), then the Model Interface Layer also 285 needs to implement re-gridding capabilities. An example is HEMCO 3.0's implementation within the CESM model, described in Sect. 3.5.

2.4 HEMCO grid

Figure 3. Benefit of using a HEMCO grid finer than the model grid for masking of emissions by a regional inventory. The example in the Figure illustrates a case where a US inventory (orange) is used to replace a global default inventory (blue) in a model simulation on a coarse 4°x5° grid. The mask is by necessity binary, so that US inventory emissions are applied only to grid cells that are 100% in the US. Using the 4×finer (2°x2.5°) HEMCO grid allows for better resolution of the US-Mexico border and hence for more of the US inventory to be used near the border. ~~This means that the emissions in the model grid cell which straddles the border will be a mix of emissions from both sides, rather than being uniquely assigned.~~

HEMCO 3.0 provides the ability for computing emissions on-line on a horizontal grid finer than the model grid. Previous versions of HEMCO assumed its operation on the same grid as the model. If the model operated multiple grids simultaneously at runtime, as is the case in WRF-GC, multiple instances of HEMCO were used, thus increasing computational and memory cost.

In HEMCO 3.0, a single instance of HEMCO reads and processes data on a user-specified HEMCO grid. When data are requested by the model, the Model Interface Layer re-grids emissions and other data from the HEMCO grid to the model grid. This allows HEMCO to (1) provide data to model components on different grids, (2) operate on a higher resolution for masking and scaling purposes, thus achieving greater accuracy at boundaries between different inventory domains, and (3) use high-resolution environmental data sets when computing emissions through extensions.

Figure 3 illustrates the benefit of using a finer HEMCO grid at the boundaries between inventories. When the HEMCO grid is disabled, HEMCO runs at model resolution, and all input data, including masks, are re-gridded by the Data Input Layer to the model resolution before emissions are computed by the HEMCO Core. In the example of Fig. 3, where a national inventory for the US is to overwrite a global default inventory, this overwriting can be done only for grid cells that are fully in the US. Grid cells straddling the border must retain the global default in order to avoid under- or over-accounting, but this then loses

information from the US inventory. This is not a problem if the national inventory straddles the border and includes information on the fractional contributions from the neighboring country, but such is not the case here. Using a finer, intermediate-resolution grid – the HEMCO grid – allows emissions at the model grid scale to more accurately blend the contributions from the two sides of the border in a single model grid cell. This also ensures greater consistency when using the same model simulations at different resolutions. As long as the HEMCO grid is kept at a single resolution, the calculated emissions will be consistent between simulations – no matter what model grid resolution is selected.

Another advantage of using a finer HEMCO grid is for emissions computed with extensions and dependent on both the model variables (provided on the model grid) and environmental data (provided on the HEMCO grid). If there is non-linear dependence of emissions on the environmental data variables, then a finer HEMCO grid will produce more accurate emissions. This is the case for example in dust emission algorithms that use land type as a categorical variable.

There is a limit to the resolution of the HEMCO grid because of the need for HEMCO to store the different inventories in memory, re-grid the data to the HEMCO grid, and process the data at higher resolution, which may be computationally expensive, even for data which may not need scaling and. In global simulations using GEOS-Chem Classic on a single machine at resolutions of 4°x5° or 2°x2.5° we have found that a HEMCO grid of 1°x1.25° is a practical limit. However, masking on the native inventory grid at much higher resolution may be desirable for regional modeling applications. One can circumvent the problem by pre-processing the emissions on their native grids using HEMCO in off-line (standalone) mode, but this adds an additional step in running the model.mode. Another option is to use a regional rather than global model. For example, in nested GEOS-Chem Classic simulations at 0.25°x0.3125° resolution we find that HEMCO can easily handle a HEMCO grid of 0.1°x0.1°, typical of the native resolution of inventories.

While any unstructured grid may be used as the HEMCO grid, it may be desirable to use a rectilinear latitude-longitude grid for prototyping HEMCO in new models. This is because the default Data Input Layer provided with HEMCO only supports rectilinear latitude-longitude grids, and most input data available in the HEMCO database library are also on rectilinear latitude-longitude grids. By choosing such a grid, the default Data Input Layer can be readily used for quick prototyping of a new HEMCO implementation, which may then be improved upon if another HEMCO grid is more desirable. In cases where other grids are used as the HEMCO grid or the model grid, conservative re-gridding needs to be implemented by the Data Input Layer or the Model Interface Layer. Examples of these scenarios are described in Sect. 3.2, where the HEMCO grid is a cubed-sphere grid and *ExtData* from MAPL with re-gridding capability is implemented as a Data Input Layer, and Sect. 3.5, where the model grid is an arbitrary grid described by an ESMF mesh file provided by the model to HEMCO, and ESMF on-line re-gridding is implemented in the Model Interface Layer.

2.5 Data broker functionality

345 HEMCO 3.0 has the capability to process any model input data other than emissions such as meteorological fields, land use maps, boundary conditions, etc. These data can be selected, subsetted (masked), added, and scaled in the same way as emissions. GEOS-Chem has long used this general data broker functionality in HEMCO, but this was previously done by interfacing directly with HEMCO's internal data containers. As this approach bypassed the HEMCO Core, processing of data by the HEMCO Core was also not supported. HEMCO 3.0 standardizes the code for models to retrieve arbitrary data from

350 HEMCO through the Model Interface Layer, thus processing all data from the Data Input Layer through the HEMCO Core. In this manner, HEMCO 3.0 can serve as a general data broker for models if desired.

3 Implementations of HEMCO 3.0 in different models

HEMCO 3.0 has been implemented so far in a number of models: GEOS-Chem 'Classic', GCHP, NASA GEOS, WRF-GC,

355 CESM2, and NOAA GEFS-Aerosol. It is planned for implementation in the NOAA UFS. These models have different architectures and software engineering environments, requiring different formulations of the Data Input Layer and Model Interface Layer with the same HEMCO Core. We describe below the particularities of implementation for each model as a guide for implementation in other models.

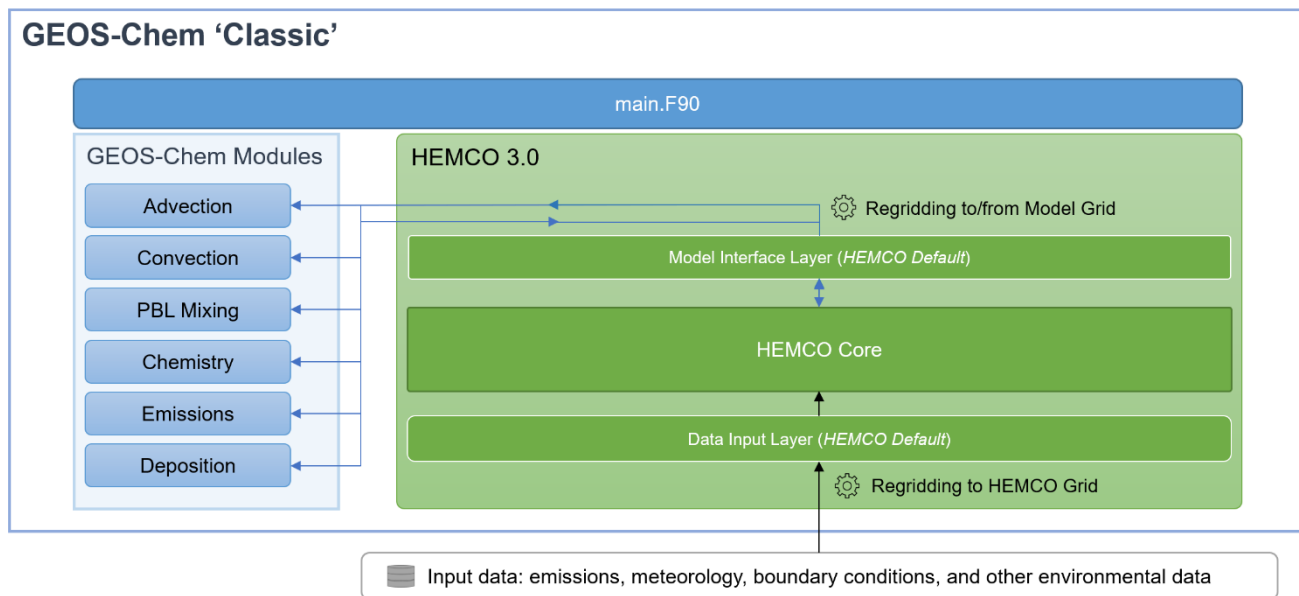
3.1 GEOS-Chem 'Classic'

360

Figure 4. HEMCO 3.0 implementation in the GEOS-Chem 'Classic' model. HEMCO in GEOS-Chem 'Classic' is a general input data broker. The main model driver routine *main.F90* successively calls each GEOS-Chem module and HEMCO in a timestep loop. HEMCO

3.0 reads all input data through the default Data Input Layer, processes the data through the HEMCO Core, and exports the data to each GEOS-Chem module through the HEMCO default Model Interface Layer. The Model Interface Layer also receives data from GEOS-Chem

365   modules that are used by HEMCO to compute state-dependent emissions through extensions.

GEOS-Chem 'Classic' (Bey et al., 2001) is an off-line chemical transport model (CTM) driven by NASA GEOS meteorological data. It operates on global or regional (nested) rectilinear latitude-longitude grids. It uses OpenMP shared-memory parallelization on a single node without a dedicated coupler. Figure 4 shows the implementation of HEMCO 3.0 in

370   GEOS-Chem 'Classic' as both an emissions component and a general input data broker. The default Data Input Layer is used to read and re-grid all input data, which are then processed through the HEMCO Core. When the HEMCO grid is different than the model grid (Sect. 2.4), the Model Interface Layer performs horizontal re-gridding between the two grids for the data flowing through it.

375   HEMCO 3.0 in GEOS-Chem 'Classic' is used for all gridded input data including not only emissions but also meteorological fields, chemical boundary conditions (for regional runs), initial conditions, and other environmental data sets such as land type, leaf area index, and sea surface salinity. It serves as a re-gridding and subsetting tool for these data. This has in particular enabled the FlexGrid capability in GEOS-Chem where regional nested domains are selected at runtime and all input data are processed for these domains (Li et al., 2021; Shen et al., 2021).

380

3.2 GEOS-Chem high performance (GCHP)

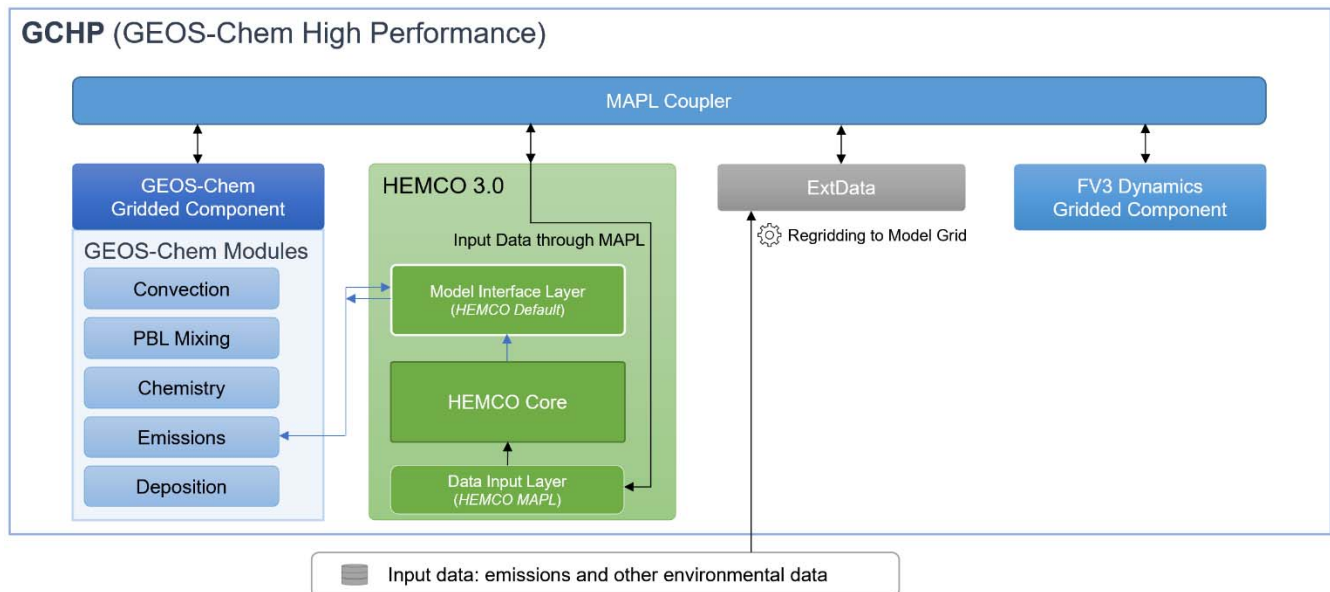



Figure 5. HEMCO 3.0 implementation in the GEOS-Chem high performance model (GCHP). GCHP operates under the MAPL coupling framework in a MPI parallel environment. The ExtData component reads all data into MAPL, re-gridding them to the model grid, and the input data are retrieved by the HEMCO MAPL implementation of the Data Input Layer. After emissions are processed by the HEMCO Core, they are exported to GEOS-Chem through the HEMCO default Model Interface Layer.

GCHP (Eastham et al., 2018; Zhuang et al., 2020) is a high-performance version of GEOS-Chem that takes advantage of the grid-independent structure of the model (Long et al., 2015) to apply a distributed-memory MPI parallelization enabling efficient simulations with thousands of cores. Implementation of MPI is through the GEOS MAPL environment on cubed-sphere grids. MAPL is a modeling toolkit built upon ESMF that provides additional tools for interfacing between ESMF and the model code (Suarez et al., 2007). It serves as a coupler for individual model components, referred to as "gridded components", and provides input and cubed-sphere re-gridding capabilities for all external data through the *ExtData* component. GCHP advection is computed by the FV3 dynamical core gridded component (Putman and Lin, 2007).

Figure 5 shows the implementation of HEMCO 3.0 in GCHP. In the MAPL environment, all data are read and re-gridded to the model grid by the *ExtData* component. Thus, HEMCO in MAPL uses the MAPL Data Input Layer, which simply retrieves data from *ExtData*. Unlike in GEOS-Chem 'Classic', meteorological data are not processed by HEMCO in GCHP, as these data are provided to GEOS-Chem through *ExtData*. There is also no option for HEMCO to operate on a HEMCO grid different from the model grid because *ExtData* re-grids all data to the model grid.

After emissions data are processed by the HEMCO Core, GEOS-Chem receives the emissions from HEMCO through the HEMCO default Model Interface Layer, in the same manner as GEOS-Chem 'Classic'. In this manner, the interface between the GEOS-Chem emissions module and HEMCO is the same for GEOS-Chem 'Classic' and GCHP, facilitating the maintenance of a single GEOS-Chem code.

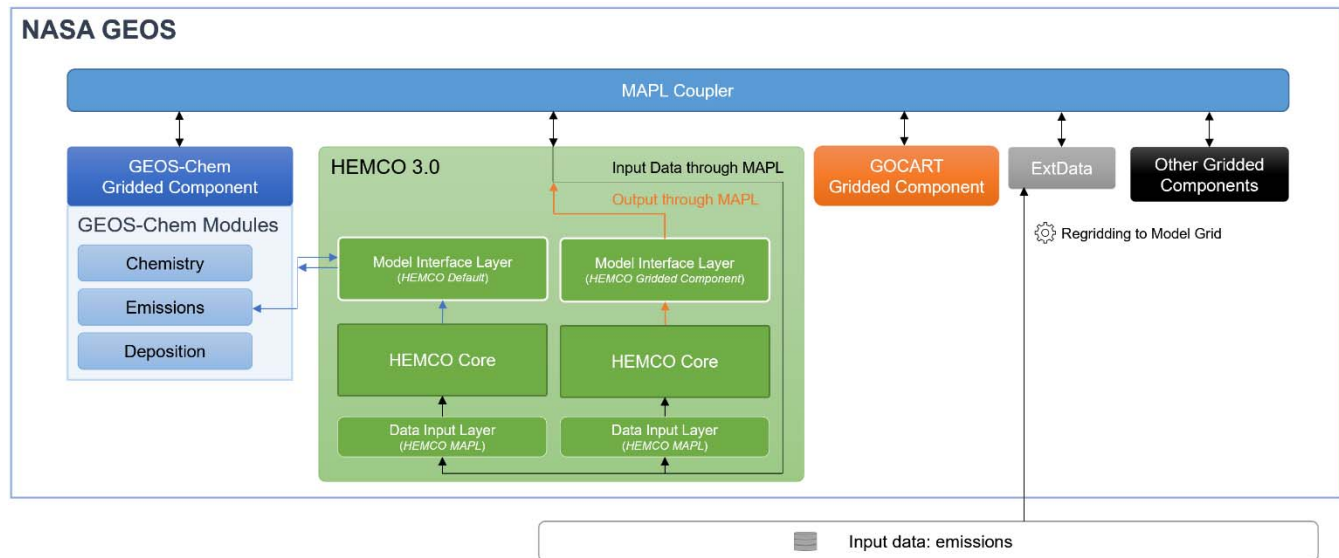



Figure 6. HEMCO 3.0 implementation in the NASA GEOS Earth System Model. GEOS is driven by the MAPL framework in a MPI parallel environment and is composed of a number of gridded components for the different ESM operators. GOCART provides a fast aerosol simulation and GEOS-Chem provides detailed chemistry. HEMCO serves both GOCART and GEOS-Chem. The ExtData component reads all data into MAPL, re-gridding them to the model grid, and the input data are retrieved by the HEMCO MAPL implementation of the Data Input Layer. GEOS-Chem and GOCART each have their own HEMCO instances. After emissions are processed by the HEMCO Core, they are exported to GEOS-Chem through the HEMCO default Model Interface Layer, and to the GOCART gridded component through a HEMCO Gridded Component Model Interface Layer.

410

415

The GEOS ESM (Rienecker et al., 2008) provides the platform for Earth system data analysis at NASA through the GEOS Data Assimilation System (GEOS-DAS). It has several ~~modules~~options for on-line representation of atmospheric chemistry including GEOS-Chem (Hu et al., 2018) and GOCART aerosols (Chin et al., 2002; Randles et al., 2017). GOCART is used in the operational GEOS-DAS as a fast module for aerosol data assimilation. The GEOS-Chem module is used in the GEOS

420 chemical forecast product (GEOS-CF; Keller et al., 2020) and in research applications. It has exactly the same code as the off-line GEOS-Chem but with all transport routines disabled, since chemical transport is done as part of the GEOS ESM atmospheric dynamics.

Figure 6 shows the implementation of HEMCO 3.0 in the GEOS ESM to serve both the GEOS-Chem and GOCART modules.

425 HEMCO 3.0 receives data from the MAPL Data Input Layer, using *ExtData* to read in data and re-gridding to the cubed-sphere grid in the GEOS ESM. The data passed to HEMCO through *ExtData* are limited to emissions, as all other data are generated through the GEOS model or introduced elsewhere.

The GEOS ESM uses separate instances of HEMCO for interfacing with GEOS-Chem and GOCART. These separate instances
430 use different HEMCO configuration files and run independently of each other in parallel. For interfacing with GEOS-Chem, HEMCO uses the Default Model Interface Layer in the same manner as GCHP (Sect. 3.2). This enables usage of GEOS-Chem chemistry within GEOS with minimal changes to the GEOS-Chem code. For interfacing with GOCART, HEMCO exports data using MAPL through a separate HEMCO Gridded Component Model Interface Layer, and those data are then imported by GOCART.

435

3.4 WRF-GC

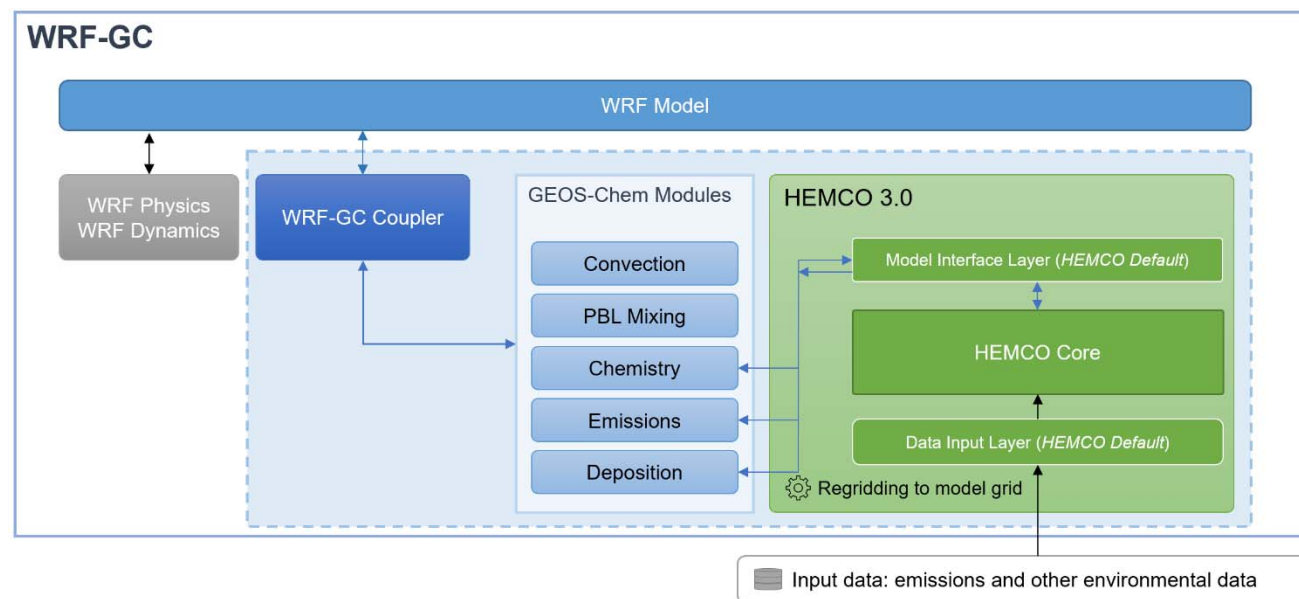



Figure 7. HEMCO 3.0 implementation in the WRF-GC model. The model is driven by WRF, which calls GEOS-Chem chemistry through the WRF-GC Coupler. HEMCO 3.0 is implemented entirely as part of GEOS-Chem in WRF-GC, with each successive domain in the WRF
440 model containing a separate instance of GEOS-Chem and HEMCO (dashed blue box). HEMCO interfaces with GEOS-Chem using the HEMCO Default Model Interface Layer and reads data through the HEMCO Default Data Input Layer, using the same approach as GEOS-Chem 'Classic'. Convection and PBL mixing are done by GEOS-Chem modules but the driving meteorological data are provided by the WRF-GC coupler instead of through HEMCO, unlike in GEOS-Chem 'Classic'.

The WRF-GC on-line model (Weather Research and Forecasting Model with GEOS-Chem chemistry; Lin et al., 2020; Feng
445 et al., 2021) couples the Weather Research and Forecasting (WRF) weather model (Skamarock et al., 2008) with GEOS-Chem, in the same manner as the coupling of WRF with WRF-Chem (Grell et al., 2005; Fast et al., 2006). It uses a WRF-GC coupler separate from the WRF and GEOS-Chem parent models to interface between the two models, converting the state between WRF and GEOS-Chem as necessary to drive both models. This coupling structure enables independent updates of each model in WRF-GC. Chemical advection is done by WRF but convection and planetary boundary layer (PBL) mixing are done by
450 GEOS-Chem using input data from WRF, following the practice in WRF-Chem. Aerosol effects on WRF radiation and cloud

physics are treated by passing GEOS-Chem aerosol information to WRF through the WRF-GC coupler. Multi-scale WRF grids communicating by 2-way nesting are also supported by WRF-GC.

Figure 7 shows the implementation structure of HEMCO within the WRF-GC model. In WRF-GC, HEMCO is implemented as a component within GEOS-Chem, using the same Default Data Input Layer and Default Model Interface Layer as GEOS-Chem 'Classic'. The Data Input Layer reads emissions and environmental data to serve GEOS-Chem emissions, chemistry, and dry deposition routines. Meteorological data simulated by WRF (including convective air mass fluxes, PBL mixing parameters, and any meteorological data needed for HEMCO extensions) are passed to GEOS-Chem by the WRF-GC coupler independently of HEMCO. Chemical initial and boundary conditions are also read by the WRF model and passed to GEOS-Chem through the WRF-GC coupler. If there are successive instances of WRF for nested domain functionality, corresponding instances of GEOS-Chem and HEMCO for each domain are used. Each instance of HEMCO operates on the grid of the corresponding WRF domain, and no separate HEMCO grid is used.

3.5 Community Earth System Model Version 2 (CESM2)

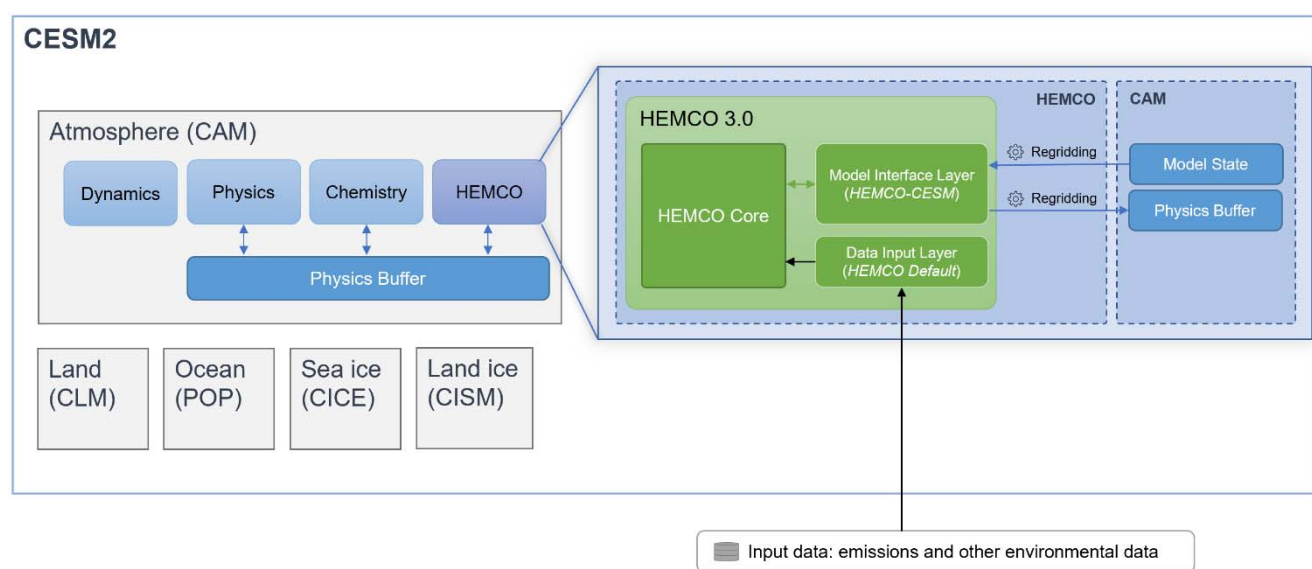



Figure 8. HEMCO 3.0 implementation in the CESM2 model. HEMCO 3.0 lives as a component within the CESM2 atmosphere (CAM), reading data through the HEMCO Default Data Input Layer, and communicating with other CESM components through the "HEMCO-CESM" implementation of the Model Interface Layer, which reads necessary meteorological quantities for computation of state-dependent emissions within HEMCO, and exports emissions and other input data to the CAM physics buffer. HEMCO operates on its own high-resolution grid, with re-gridding routines to pass data to/from CAM. Environmental data for computing emissions and dry deposition (e.g., land type) may either be read from HEMCO data libraries or provided to HEMCO by the CAM state.

The Community Earth System Model Version 2 (CESM2) is an open-source model enabling a wide range of Earth science simulations including atmospheric chemistry. The atmospheric component of CESM2 is the Community Atmosphere Model

(CAM), including the CAM-chem module to simulate chemistry (Emmons et al., 2020). The new MUSICA initiative at NCAR (Pfister et al., 2020) seeks to expand the capabilities and versatility of the chemical simulation within CAM, including use of GEOS-Chem as an alternative module.

475

Figure 8 shows the implementation of HEMCO 3.0 in CESM2. HEMCO serves emissions to CAM-chem, GEOS-Chem, and potentially to any representation of atmospheric chemistry in CAM. The Model Interface Layer includes routines to export data processed by HEMCO to CAM's physics buffer, a temporary storage space for model components to share data at runtime. As CAM supports a variety of grids, during initialization, the HEMCO-CESM Model Interface Layer reads the ESMF mesh file that describes the grid used by CAM, and uses ESMF on-line re-gridding (https://earthsystemmodeling.org/regrid/) to re-grid data between HEMCO and CAM. By using ESMF re-gridding capabilities, HEMCO is capable of operating independently of the grid used by the model, as long as the model grid description is provided to HEMCO using an ESMF mesh file.

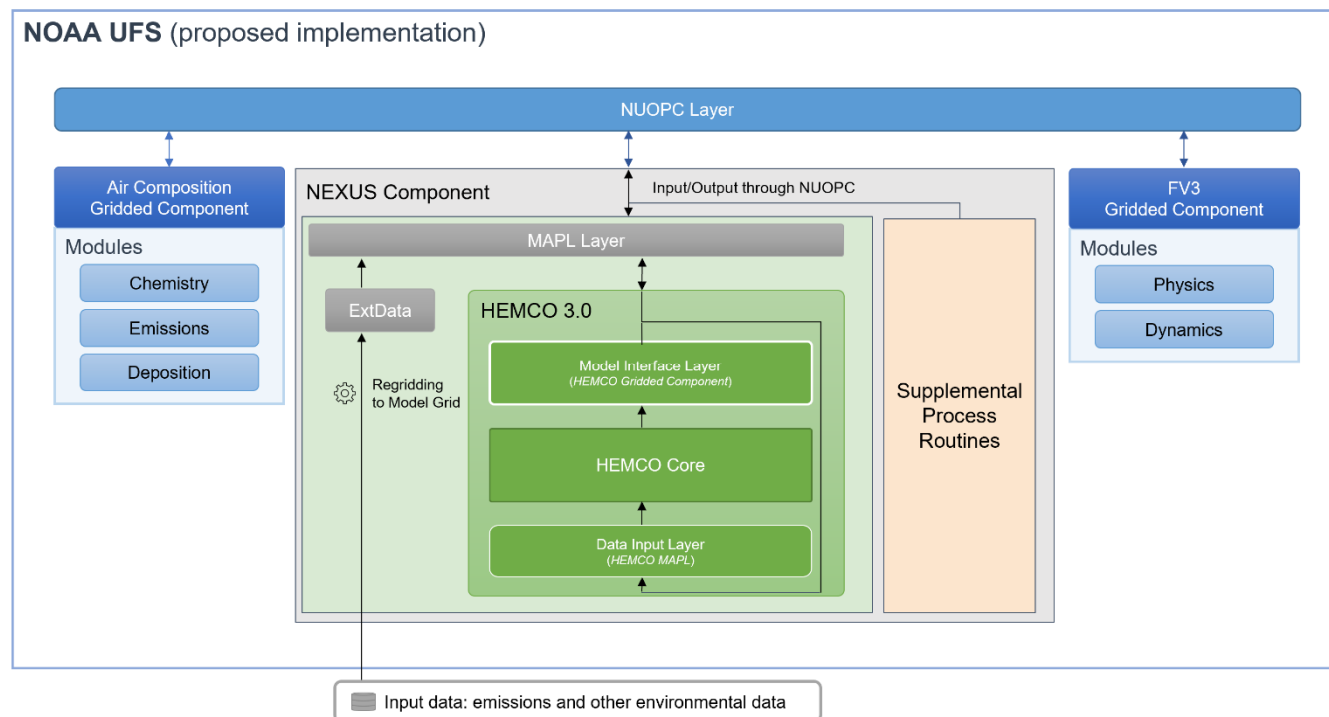
480

485 HEMCO within CESM2 can use the existing HEMCO emissions database library out-of-the-box and it can also import new emission inventories and extensions. The CAM-chem implementation uses a configuration file with the appropriate CAM-chem species mapping as described by Emmons et al. (2020). An example for alcohols was described in Sect. 2.1. For CESM with GEOS-Chem chemistry (CESM2-GC), HEMCO works out-of-the-box with configuration files from GEOS-Chem since CESM2-GC uses the same species.

490

Environmental data needed for computing emissions and dry deposition, such as land type and leaf area index, are normally provided by the CESM state to HEMCO through the Model Interface Layer. This is required for coupled chemistry-biosphere-climate simulations, where atmospheric chemistry affects ecosystem state (both directly and indirectly through the climate), which in turns affects atmospheric chemistry. Alternatively, one may want to use independent environmental data specified

495 through the HEMCO configuration file in order to compare the CESM simulation to an independent simulation of atmospheric chemistry such as with GEOS-Chem 'Classic'. Both capabilities are supported by HEMCO within CESM2.

3.6 NOAA GEFS-Aerosol and NOAA UFS

500 **Figure 8. Planned HEMCO 3.0 implementation in the NOAA UFS.** UFS is a unified software infrastructure for NOAA forecast models to enable exchange of model components between different application models. HEMCO is planned as the core of the NOAA Emissions and eXchange Unified System (NEXUS) component to serve both emissions and surface exchange (state-dependent emissions, dry deposition, and two-way exchange) as part of UFS. HEMCO operates under the MAPL coupling framework in the same way as in the NASA GEOS ESM (Fig. 6) but reads independent environmental data for use in surface exchange calculations. Supplemental process

505 routines provide the surface exchange calculations ~~supplemental to HEMCO.~~. NEXUS operates under the NUOPC coupling framework to provide emissions and surface exchange information to other UFS components.

As described in NOAA's 2018 Strategic Implementation Plan for next-generation modeling systems (https://www.weather.gov/sti/stimodeling_nggps_implementation), "a unified emission system with the capability of providing model-ready, global anthropogenic and natural source emissions inputs for aerosol and gas phase atmospheric

510 composition across scales is needed." The next generation modeling systems of NOAA are being realized as the Unified Forecast System (UFS; https://ufscommunity.org), which will replace the current suite of forecast models in the National Weather Service (NWS) over the next few years. The UFS does not refer to a single modeling system, but rather describes a unified software infrastructure that permits the exchange of model components between different application models. To respond to the requirement for a unified emissions system, HEMCO 3.0 is being tested to serve as the core of the NOAA

515 Emissions and eXchange Unified System (NEXUS) component (Campbell et al., 2020). NEXUS will provide emissions and broader surface exchange information for NOAA's UFS global and regional aerosol and atmospheric composition (AAC) models.

Figure 8 shows the planned implementation of HEMCO as the core of NEXUS within the NOAA UFS. Currently, HEMCO is used as an offline emissions pre-processor for an experimental version of the Global Ensemble Forecast System-Aerosol (GEFS-Aerosol). GEFS-Aerosol is a global aerosol model based on the Global Forecast System (GFS), which is based on the Finite Volume Cubed-Sphere (FV3) dynamical core (Lin et al., 1994; Lin and Rood, 1996; Lin, 2004; Putman and Lin, 2007; Chen et al., 2013; Harris and Lin, 2013; Harris et al., 2016; Zhou et al., 2019). The operational version of GEFS-Aerosol is run by the NWS as a special unperturbed forecast of the Global Ensemble Forecast System version 12, which provides an ensemble forecast product four times per day.

Ongoing development of NEXUS includes adaptation to provide emissions for three future UFS applications: i) a global aerosol model, adapted from GEFS-Aerosol, that will be part of a sub-seasonal to seasonal forecast capability, and which uses the NASA GOCART aerosol model; ii) an on-line, regional-scale air quality model with fully coupled gas-aerosol chemistry derived from the Community Multiscale Air Quality (CMAQ) model (U. S. EPA, 2018); and, iii) an on-line, regional-scale rapid refresh forecast system (RRFS) that includes fully coupled smoke and dust emissions and transport and aerosol-weather interactions. Further planned development of NEXUS for UFS AAC models includes the integration of HEMCO 3.0 as a dynamic, on-line emissions processor for both anthropogenic inventories as well as natural, process-based sources (e.g., windblown dust, wildfire smoke, sea-salt, etc.) ~~As the UFS AAC models evolve over time, expanding global chemical simulation capabilities and including a broader suite of chemical/aerosol options in the regional models, NEXUS will be fully capable of providing emissions for both research and operational applications.~~ A longer-term goal for NEXUS includes the harmonization of emissions-related processes with the surface-atmosphere exchange and boundary-layer processes in the land surface modeling system. The current vision for the NEXUS architecture is evolving as the UFS AAC models are being developed but will rely on established coupling and integration infrastructures, such as the ~~the~~ National Unified Operational Prediction Capability (NUOPC) Layer (Theurich et al., 2016; http://earthsystemmodeling.org/nuopc/).

4 Conclusions

We presented an updated version 3.0 of the Harmonized Emissions Component (HEMCO 3.0) for atmospheric models. HEMCO is a versatile on-line emissions processor originally developed for the GEOS-Chem chemical transport model but is now portable to any atmospheric model. HEMCO allows users to select an ensemble of emission inventories and state-dependent emission algorithms (extensions) with capabilities for re-gridding, adding, masking, and scaling emission data and mapping them to model species. HEMCO 3.0 addresses limitations of the previous version, HEMCO 2.0, which used features specific to the GEOS-Chem 'Classic' or MAPL environments, and was limited to operating on the model grid. HEMCO 3.0 has a modular structure to facilitate its implementation in models with different software engineering protocols. It features an optional high-resolution grid that may be finer than the model grid for more accurate masking, more accurate computation of emissions with nonlinear algorithms, and the serving of emission data to multi-grid models with greater computational and

memory efficiency. HEMCO 3.0 can also serve as a general data broker to process all input data in the atmospheric model, not just emissions.

HEMCO 3.0 modularizes the original HEMCO (Keller et al., 2014) into three layers: the Data Input layer, the HEMCO Core, and the Model Interface Layer. The Data Input Layer reads a configuration file that defines the emission environment desired by the user, extracts the necessary inventory and other data files in netCDF format from the HEMCO database library, and re-grids the data to the model grid or to the higher-resolution HEMCO grid. The HEMCO Core subsets, adds, masks, and scales the different data sets as specified by the configuration file. The Model Interface Layer collects the emissions data from the HEMCO Core to pass on to the atmospheric model, and also passes model state variables to the HEMCO Core for computing emissions through extensions. The HEMCO Core and database library are common to HEMCO implementations across all models. The Data Input Layer and Model Interface Layer may be used out-of-the-box or modified to fit a model's specific architecture.

HEMCO 3.0 has been implemented in several models: the GEOS-Chem CTM in both 'Classic' and High-Performance (GCHP) configurations, the NASA GEOS ESM, WRF with GEOS-Chem chemistry (WRF-GC), the CESM2 model with either CAM-chem or GEOS-Chem chemistry, and the NOAA GEFS-Aerosol model as an offline emissions pre-processor. GEOS-Chem 'Classic' relies on the default implementations of the Data Input Layer and Model Interface Layer, and these defaults may be used for quick implementation of HEMCO 3.0 in any model. GCHP and the GEOS ESM use the MAPL coupler built on ESMF to read and re-grid data; in that case the corresponding functionalities are removed from the HEMCO Data Input Layer with no editing of code in the HEMCO Core. HEMCO 3.0 is planned for inclusion in the NOAA UFS as the core of the NEXUS component that will serve emission and surface exchange information to the suite of NOAA aerosol and atmospheric composition forecast models. This will add a new dimension to HEMCO capabilities to include surface deposition and two-way exchange of chemical species.

Implementation of HEMCO 3.0 in the CESM2 model is an important step in the development of MUSICA, a flexible modeling framework for the next-generation CESM allowing for versatile use of different atmospheric chemistry simulation components on any grid and scale (Pfister et al., 2020). HEMCO within CESM can operate on multi-scale or unstructured grids, can serve data to any CESM atmospheric component, and can interface with any chemical mechanism by mapping emitted species to the mechanism species. Through HEMCO, CESM users can readily use and combine any ensemble of emission inventory data and algorithms that they choose, independently of their chemical mechanism or other aspects of the chemical simulation. HEMCO thus provides a general vessel for the treatment of emissions in MUSICA, and could also in the future provide a general input data broker facility for CESM.

585 *Code availability.* The code used in this paper is permanently archived at https://github.com/jimmielin/HEMCO3-Paper-Code (https://doi.org/10.5281/zenodo.4706173).

Author contributions. DJJ, SDE, LKE provided project oversight. HL, EWL, MPS, CAK developed the HEMCO 3.0 code. HL, EWL, MPS, CAK developed the HEMCO to GEOS-Chem 'Classic' interface. EWL, CAK developed the HEMCO MAPL

590 interface for use in GCHP. CAK developed the MAPL gridded component for use in GEOS. HL, TF, SDE and LKE designed and developed the HEMCO-CESM interface. PCC, BB, RDS and RM designed the NOAA UFS interface to HEMCO. HL wrote the paper with assistance from DJJ and input from all co-authors.

Competing interests. The authors declare no competing interests.

595

Acknowledgements. This project was supported by the Atmospheric Chemistry Program of the US National Science Foundation and by the NASA Atmospheric Composition Modeling and Analysis Program. The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977.

600 Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. The authors would also like to thank Steve Goldhaber and Andrew Conley for useful discussions in designing and developing the HEMCO-CESM interface.

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, Atmos. Chem. Phys., 11, 4039–4072, https://doi.org/10.5194/acp-11-4039-2011, 2011.

Andela, N., Kaiser, J., Heil, A., van Leeuwen, T. T., Wooster, M. J., van der Werf, G.R., Remy, S., and Schultz, M. G.: Assessment of the Global Fire Assimilation System (GFASv1), ECMWF Technical Memoranda, 702, 1-70, https://doi.org/10.21957/7pg36pe5m, 2013.

Andreae, M. O., and Merlet, P.: Emission of trace gases and aerosols from biomass burning, Global Biogeochem. Cy., 15(4), 955-966, https://doi.org/10.1029/2000GB001382, 2001.

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, J. Geophys. Res. Atmos., 106, 23 073–23 095, https://doi.org/10.1029/2001JD000807, 2001.

Bouwman, A. F., Lee, D. S., Asman, W. A. H., Dentener, F. J., Van Der Hoek, K. W., and Olivier, J. G. J.: A global high-resolution emission inventory for ammonia, Global Biogeochem. Cy., 11(4), 561-587, https://doi.org/10.1029/97GB02266, 1997.

Byun, D., and Schere, K. L.: Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, Appl. Mech. Rev., 59(2), 51-77, https://doi.org/10.1115/1.2128636, 2006.

Campbell, P., Baker, B., Saylor, R., Tong, D., Tang, Y., Lee, P., McKeen, S., Frost, G., and Keller, C.: Initial Development of a NOAA Emissions and eXchange Unified System (NEXUS), 100th American Meteorological Society Conference, Boston, MA, https://doi.org/10.13140/RG.2.2.21070.20806, 2020.

Carn, S. A., Yang, K., Prata, A. J., and Krotkov, N. A.: Extending the long-term record of volcanic SO2 emissions with the Ozone Mapping and Profiler Suite nadir mapper, Geophys. Res. Lett., 42(3), 925-932, https://doi.org/10.1002/2014GL062437, 2015.

Carpenter, L. J., MacDonald, S. M., Shaw, M. D., Kumar, R., Saunders, R. W., Parthipan, R., Wilson, J., and Plane, J. M. C.: Atmospheric iodine levels influenced by sea surface emissions of inorganic iodine, Nat. Geosci., 6, 108-111, https://doi.org/10.1038/ngeo1687, 2013.

Chen, X., Andronova, N., Van Leer, B., Penner, J. E., Boyd, J. P., Jablonowski, C., and Lin, S.-J.: A Control-Volume Model of the Compressible Euler Equations with a Vertical Lagrangian Coordinate, Monthly Weather Review, 141(7), 2526-2544, https://doi.org/10.1175/MWR-D-12-00129.1, 2013.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun

Photometer Measurements, J. Atmos. Sci., 59(3), 461-483, https://doi.org/10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2, 2002.

Corbett, J. J., Fischbeck, P. S., and Pandis, S. N.: Global nitrogen and sulfur inventories for oceangoing ships, J. Geophys. Res. Atmos., 104(D3), 3457-3470, https://doi.org/10.1029/1998JD100040, 1999.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., Monni, S., Doering, U., Olivier, J. G. J., Pagliari, V., and Janssens-Maenhout, G.: Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2, Earth Syst. Sci. Data, 10, 1987–2013, https://doi.org/10.5194/essd-10-1987-2018, 2018.

Croft, B., Wentworth, G. R., Martin, R. V., Leaitch, W. R., Murphy, J. G., Murphy, B. N., Kodros, J. K., Abbatt, J. P. D., and Pierce, J. R.: Contribution of Arctic seabird-colony ammonia to atmospheric particles and cloud-albedo radiative effect, Nat. Commun., 7, 13444, https://doi.org/10.1038/ncomms13444, 2016.

Darmenov, A., and da Silva, A. M.: The Quick Fire Emissions Dataset (QFED) - Documentation of versions 2.1, 2.2 and 2.4, NASA Technical Memorandum, 38, NASA/TM-2015-104606, https://ntrs.nasa.gov/api/citations/20180005253/downloads/20180005253.pdf, 2015.

Di Giuseppe, F., Rémy, S., Pappenberger, F., and Wetterhall, F.: Improving GFAS and CAMS biomass burning estimations by means of the Global ECMWF Fire Forecast system (GEFF), ECMWF Technical Memoranda, 790, 1-20, https://doi.org/10.21957/uygqqtyh7, 2016.

Di Giuseppe, F., Rémy, S., Pappenberger, F., and Wetterhall, F.: Using the Fire Weather Index (FWI) to improve the estimation of fire emissions from fire radiative power (FRP) observations, Atmos. Chem. Phys., 18, 5359–5370, https://doi.org/10.5194/acp-18-5359-2018, 2018.

Eastham, S. D., Long, M. S., Keller, C. A., Lundgren, E., Yantosca, R. M., Zhuang, J., Li, C., Lee, C. J., Yannetti, M., Auer, B. M., Clune, T. L., Kouatchou, J., Putman, W. M., Thompson, M. A., Trayanov, A. L., Molod, A. M., Martin, R. V., and Jacob, D. J.: GEOS-Chem High Performance (GCHP v11-02c): a next-generation implementation of the GEOS-Chem chemical transport model for massively parallel applications, Geosci. Model. Dev., 11, 2941–2953, https://doi.org/10.5194/gmd-11-2941-2018, 2018.

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), J. Adv. Model Earth Sy., 12, e2019MS001882, https://doi.org/10.1029/2019MS001882, 2020.

Eyring, V., Köhler, H. W., van Aardenne, J., and Lauer, A.: Emissions from international shipping: 1. The last 50 years, J. Geophys. Res. Atmos., 110, D17305, https://doi.org/10.1029/2004JD005619, 2005.

Fairlie, T. D., Jacob, D. J., Dibb, J. E., Alexander, B., Avery, M. A., van Donkelaar, A., and Zhang, L.: Impact of mineral dust on nitrate, sulfate, and ozone in transpacific Asian pollution plumes, Atmos. Chem. Phys., 10, 3999–4012, https://doi.org/10.5194/acp-10-3999-2010, 2010.

670 Fast, J. D., Gustafson Jr., W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., Grell, G. A., and Peckham, S. E.: Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model, J. Geophys. Res.-Atmos., 111, D21305, https://doi.org/10.1029/2005JD006721, 2006.

Feng, X., Lin, H., Fu, T.-M., Sulprizio, M. P., Zhuang, J., Jacob, D. J., Tian, H., Ma, Y., Zhang, L., Wang, X., ~~and~~ Chen, Q.~~.~~,
675 and Han, Z.: WRF-GC (v2.0): online two-way coupling of WRF (v3.9.1.1) and GEOS-Chem (v12.7.2) for modeling regional atmospheric chemistry~~-~~ meteorology interactions, Geosci. Model Dev.~~ Discuss.,~~., 14, 3741–3768, https://doi.org/10.5194/gmd-~~2020-441, in review,~~14-3741-2021~~.~~

, 2021.Freitas, S. R., Longo, K. M., Alonso, M. F., Pirre, M., Marecal, V., Grell, G., Stockler, R., Mello, R. F., and Sánchez Gácita, M.: PREP-CHEM-SRC – 1.0: a preprocessor of trace gas and aerosol emission fields for regional and global
680 atmospheric chemistry models, Geosci. Model Dev., 4, 419–433, https://doi.org/10.5194/gmd-4-419-2011, 2011.

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

Ge, C., Wang, J., Carn, S., Yang, K., Ginoux, P., and Krotkov, N.: Satellite-based global volcanic SO2 emissions and sulfate
685 direct radiative forcing during 2005–2012, J. Geophys. Res. Atmos., 121(7), 3446-3464, https://doi.org/10.1002/2015JD023134, 2016.

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), J. Geophys. Res. Biogeo., 118(1), 317-328, https://doi.org/10.1002/jgrg.20042, 2013.

690 Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106(D17), 20255-20273, https://doi.org/10.1029/2000JD000053, 2001.

Gong, S. L.: A parameterization of sea-salt aerosol source function for sub- and super-micron particles, Global Biogeochem. Cy., 17(4), 1097, https://doi.org/10.1029/2003GB002079, 2003.

695 Granier, C., Lamarque, J. F., Mieville, A., Muller, J. F., Olivier, J., Orlando, J., Peters, J., Petron, G., Tyndall, G., and Wallens, S.: POET, a database of surface emissions of ozone precursors, http://www.aero.jussieu.fr/projet/ACCENT/POET.php (last retrieved: 23 Jul 2021), 2005.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, https://doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

700 Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model. Dev., 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Harris, L. M., and Lin, S.-J.: A Two-Way Nested Global-Regional Dynamical Core on the Cubed-Sphere Grid, Mon. Weather Rev., 141(1), 283-306, https://doi.org/10.1175/MWR-D-11-00201.1, 2013.

705 Harris, L. M., Lin, S.-J., and Tu, C.: High-Resolution Climate Simulations Using GFDL HiRAM with a Stretched Global Grid, J. Climate, 29(11), 4293-4314, https://doi.org/10.1175/JCLI-D-15-0389.1, 2016.

Heil, A., Kaiser, J., van der Werf, G. R., Wooster, M. J., Schultz, M. G., and van der Gon, H. D.: Assessment of the Real-Time Fire Emissions (GFASv0) by MACC, ECMWF Technical Memoranda, 628, 1-47, https://doi.org/10.21957/2m000mza9, 2010.

710 Hill, C., DeLuca, C., Balaji, V., Suarez, M., and Silva, A. D.: The Architecture of the Earth System Modeling Framework, Comput. Sci. Eng., 6, 18–28, http://doi.org/10.1109/MCISE.2004.1255817, 2004.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community

715 Emissions Data System (CEDS), Geosci. Model Dev., 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Holmes, C. D., Prather, M. J., Søvde, O. A., and Myhre, G.: Future methane, hydroxyl, and their uncertainties: key climate and emission parameters for future predictions, Atmos. Chem. Phys., 13, 285–302, https://doi.org/10.5194/acp-13-285-2013, 2013.

Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., Nielsen, J. E., Pawson, S., Thompson, M. A., Trayanov,

720 A. L., Travis, K. R., Grange, S. K., Evans, M. J., and Jacob, D. J.: Global simulation of tropospheric chemistry at 12.5 km resolution: performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM), Geosci. Model Dev., 11, 4603–4620, https://doi.org/10.5194/gmd-11-4603-2018, 2018.

Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V., Russell, A. R., Valin, L. C., and Cohen, R. C.: Steps towards a mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints, Atmos. Chem. Phys.,

725 12, 7779–7795, https://doi.org/10.5194/acp-12-7779-2012, 2012.

Jaeglé, L., Quinn, P. K., Bates, T. S., Alexander, B., and Lin, J.-T.: Global distribution of sea salt aerosols: new constraints from in situ and remote sensing observations, Atmos. Chem. Phys., 11, 3137–3157, https://doi.org/10.5194/acp-11-3137-2011, 2011.

Jähn, M., Kuhlmann, G., Mu, Q., Haussaire, J.-M., Ochsner, D., Osterried, K., Clément, V., and Brunner, D.: An online

730 emission module for atmospheric chemistry transport models: implementation in COSMO-GHG v5.6a and COSMO-ART v5.1-3.1, Geosci. Model Dev., 13, 2379-2392, https://doi.org/10.5194/gmd-13-2379-2020, 2020.

Janssens-Maenhout, G., Dentener, F., van Aardenne, J., Monni, S., Pagliari, V., Orlandini, L., Klimont, Z., Kurokawa, J., Akimoto, H., Ohara, T., Wankmüller, R., Battye, B., Grano, D., Zuber, A., and Keating, T.: EDGAR-HTAP: a harmonized gridded air pollution emission dataset based on national inventories, JRC Scientific and Technical Reports, 25229, 1-42,

735 https://doi.org/10.2788/14102, 2012.

Jöckel, P.: Technical note: Recursive rediscretisation of geo-scientific data in the Modular Earth Submodel System (MESSy), Atmos. Chem. Phys., 6, 3557–3562, https://doi.org/10.5194/acp-6-3557-2006, 2006.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527–554, https://doi.org/10.5194/bg-9-527-2012, 2012.

Keller, C. A., Long, M. S., Yantosca, R. M., Da Silva, A. M., Pawson, S., and Jacob, D. J.: HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models, Geosci. Model Dev., 7, 1409–1417, https://doi.org/10.5194/gmd-7-1409-2014, 2014.

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob, D. J., and Pawson, S.: Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0, Earth and Space Science Open Archive (preprint), 1-38, https://doi.org/10.1002/essoar.10505287.1, 2020.

Li, K., Jacob, D. J., Liao, H., Qiu, Y., Shen, L., Zhai, S., Bates, K. H., Sulprizio, M. P., Song, S., Lu, X., Zhang, Q., and Zheng, B.: Ozone pollution in the North China Plain spreading into the late-winter haze season, PNAS, 118(10), e2015797118, https://doi.org/10.1073/pnas.2015797118, 2021.

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., and He, K.: Anthropogenic emission inventories in China: a review, Natl. Sci. Rev., 4(6), 834-866, https://doi.org/10.1093/nsr/nwx150, 2017.

Liang, Q., Stolarski, R. S., Kawa, S. R., Nielsen, J. E., Douglass, A. R., Rodriguez, J. M., Blake, D. R., Atlas, E. L., and Ott, L. E.: Finding the missing stratospheric Bry: a global modeling study of CHBr3 and CH2Br2, Atmos. Chem. Phys., 10, 2269–2286, https://doi.org/10.5194/acp-10-2269-2010, 2010.

Lin, S.-J., Chao, W. C., Sud, Y. C., and Walker, G. K.: A Class of the van Leer-type Transport Schemes and Its Application to the Moisture Transport in a General Circulation Model, Mon. Weather Rev., 122(7), 1575-1593, https://doi.org/10.1175/1520-0493(1994)122<1575:ACOTVL>2.0.CO;2, 1994.

Lin, S.-J., and Rood, R. B.: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes, Mon. Weather. Rev., 124(9), 2046-2070, https://doi.org/10.1175/1520-0493(1996)124<2046:MFFSLT>2.0.CO;2, 1996.

Lin, S.-J.: A "Vertically Lagrangian" Finite-Volume Dynamical Core for Global Models, Mon. Weather Rev., 132(10), 2293-2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2, 2004.

Lin, H., Feng, X., Fu, T.-M., Tian, H., Ma, Y., Zhang, L., Jacob, D. J., Yantosca, R. M., Sulprizio, M. P., Lundgren, E. W., Zhuang, J., Zhang, Q., Lu, X., Zhang, L., Shen, L., Guo, J., Eastham, S. D., and Keller, C. A.: WRF-GC (v1.0): online coupling of WRF (v3.9.1.1) and GEOS-Chem (v12.2.1) for regional atmospheric chemistry modeling – Part 1: Description of the one-way model, Geosci. Model Dev., 13, 3241–3265, https://doi.org/10.5194/gmd-13-3241-2020, 2020.

Long, M. S., Yantosca, R., Nielsen, J. E., Keller, C. A., da Silva, A., Sulprizio, M. P., Pawson, S., and Jacob, D. J.: Development of a grid-independent GEOS-Chem chemical transport model (v9-02) as an atmospheric chemistry module
770   for Earth system models, Geosci. Model. Dev., 8, 595-602, https://doi.org/10.5194/gmd-8-595-2015, 2015.

Lu, H., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., Nielsen, J. E., Pawson, S., Thompson, M. A., Trayanov, A. L., Travis, K. R., Grange, S. K., Evans, M. J., and Jacob, D. J.: Global simulation of tropospheric chemistry at 12.5 km resolution: performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM), Geosci. Model. Dev., 11, 4603-4620, https://doi.org/10.5194/gmd-11-4603-2018, 2018.

775   MacDonald, S. M., Gómez Martín, J. C., Chance, R., Warriner, S., Saiz-Lopez, A., Carpenter, L. J., and Plane, J. M. C.: A laboratory characterisation of inorganic iodine emissions from the sea surface: dependence on oceanic variables and parameterisation for global modelling, Atmos. Chem. Phys., 14, 5841–5852, https://doi.org/10.5194/acp-14-5841-2014, 2014.

Marais, E. A., and Wiedinmyer, C.: Air Quality Impact of Diffuse and Inefficient Combustion Emissions in Africa (DICE-
780   Africa), Environ. Sci. Technol., 2016, 50(19), 10739-10745, https://doi.org/10.1021/acs.est.6b02602, 2016.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V.: A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS), Earth Syst. Sci. Data, 12, 3413–3442, https://doi.org/10.5194/essd-12-3413-2020, 2020.

785   Millet, D. B., Guenther, A., Siegel, D. A., Nelson, N. B., Singh, H. B., de Gouw, J. A., Warneke, C., Williams, J., Eerdekens, G., Sinha, V., Karl, T., Flocke, F., Apel, E., Riemer, D. D., Palmer, P. I., and Barkley, M.: Global atmospheric budget of acetaldehyde: 3-D model analysis and constraints from in-situ and satellite observations, Atmos. Chem. Phys., 10, 3405–3425, https://doi.org/10.5194/acp-10-3405-2010, 2010.

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A. L., Field, R. D., Arneth, A., Forrest, M., Hantson, S.,
790   Kehrwald, N. M., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Yue, C., Kaiser, J. W., and van der Werf, G. R.: Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015), Geosci. Model Dev., 10, 3329–3357, https://doi.org/10.5194/gmd-10-3329-2017, 2017.

Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, Geosci. Model Dev., 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015,
795   2015.

Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C., and Koshak, W. J.: Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data, J. Geophys. Res., 117, D20307, https://doi.org/10.1029/2012JD017934, 2012.

Olivier, J., Peters, J., Granier, C., Petron, G., Muller, J. F., and Wallens, S.: Present and future surface emissions of atmospheric
800   compounds, POET report #2, EU project EVK2-1999-00011, http://accent.aero.jussieu.fr/Documents/del2_final.doc (last retrieved: 23 Jul 2021), 2003.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, BioScience, 51, 933–938, https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2, 2001.

Ordóñez, C., Lamarque, J.-F., Tilmes, S., Kinnison, D. E., Atlas, E. L., Blake, D. R., Sousa Santos, G., Brasseur, G., and Saiz-Lopez, A.: Bromine and iodine chemistry in a global chemistry-climate model: description and evaluation of very short-lived oceanic sources, Atmos. Chem. Phys., 12, 1423–1447, https://doi.org/10.5194/acp-12-1423-2012, 2012.

Ott, L. E., Pickering, K. E., Stenchikov, G. L., Allen, D. J., DeCaria, A. J., Ridley, B., Lin, R.-F., Lang, S., and Tao, W.-K., Production of lightning NOx and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations, J. Geophys. Res. Atmos., 115(D4), D04301, https://doi.org/10.1029/2009JD011880, 2010.

Pfister, G. G., Eastham, S. D., Arellano, A. F., Aumont, B., Barsanti, K. C., Barth, M. C., Conley, A., Davis, N. A., Emmons, L. K., Fast, J. D., Fiore, A. M., Gaubert, B., Goldhaber, S., Granier, C., Grell, G. A., Guevara, M., Henze, D. K., Hodzic, A., Liu, X., Marsh, D. R., Orlando, J. J., Plane, J. M. C., Polvani, L. M., Rosenlof, K. H., Steiner, A. L., Jacob, D. J., and Brasseur, G. P.: The Multi-Scale Infrastructure for Chemistry and Aerosols (MUSICA), Bull. Amer. Meteor. Soc., 101(10), E1743-E1760, https://doi.org/10.1175/BAMS-D-19-0331.1, 2020.

Philip, S., Martin, R. V., Snider, G., Weagle, C. L., van Donkelaar, A., Brauer, M., Henze, D. K., Klimont, Z., Venkataraman, C., Guttikunda, S. K., and Zhang, Q.: Anthropogenic fugitive, combustion and industrial dust is a significant, underrepresented fine particulate matter source in global atmospheric models, Environ. Res. Lett., 12(4), 044018, https://doi.org/10.1088/1748-9326/aa65a4, 2017.

Putman, W. M., and Lin, S.-J.: Finite-volume transport on various cubed-sphere grids, J. Comput. Phys., 227(1), 55-78, https://doi.org/10.1016/j.jcp.2007.07.022, 2007.

Ramboll Environment and Health: Comprehensive Air Quality Model with Extensions (CAMx) User's Guide, Version 7.10, http://www.camx.com/files/camxusersguide_v7-10.pdf, 2020.

Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M., and Morton, D. C.: Global burned area and biomass burning emissions from small fires, J. Geophys. Res. Biogeo., 117, G04012, https://doi.org/10.1029/2012JG002128, 2012.

Rémy, S., Veira, A., Paugam, R., Sofiev, M., Kaiser, J. W., Marenco, F., Burton, S. P., Benedetti, A., Engelen, R. J., Ferrare, R., and Hair, J. W.: Two global data sets of daily fire emission injection heights since 2003, Atmos. Chem. Phys., 17, 2921–2942, https://doi.org/10.5194/acp-17-2921-2017, 2017.

Rienecker, M. M., Suarez, M. J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R. D., Gelaro, R., Stajner, I., and Nielsen, J. E.: The GEOS-5 Data Assimilation System-Documentation of versions 5.0.1 and 5.1.0, and 5.2.0. NASA Tech. Rep. Series on Global Modeling and Data Assimilation, NASA/TM-2008-104606, Vol. 27, 92 pp., 2008.

835 Shen, L., Zavala-Araiza, D., Gautam, R., Omara, M., Scarpelli, T., Sheng, J., Sulprizio, M.P., Zhuang, J., Zhang, Y., Qu, Z., Lu, X., Hamburg, S., and Jacob, D.J.: Unravelling a large methane emission discrepancy in Mexico using satellite observations, submitted, 2021.

Sherwen, T., Evans, M. J., Carpenter, L. J., Andrews, S. J., Lidster, R. T., Dix, B., Koenig, T. K., Sinreich, R., Ortega, I., Volkamer, R., Saiz-Lopez, A., Prados-Roman, C., Mahajan, A. S., and Ordóñez, C.: Iodine's impact on tropospheric
840 oxidants: a global model study in GEOS-Chem, Atmos. Chem. Phys., 16, 1161–1186, https://doi.org/10.5194/acp-16-1161-2016, 2016.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: NCAR Tech. Note NCAR/TN-475+STR: A Description of the Advanced Research WRF Version 3, https://doi.org/10.5065/D68S4MVH, 2008.

845 Stettler, M. E. J., Eastham, S., and Barrett, S. R. H.: Air quality and public health impacts of UK airports. Part I: Emissions, Atmos. Environ., 45(31), 5415-5424, https://doi.org/10.1016/j.atmosenv.2011.07.012, 2011.

Suarez, M., Trayanov, A., Hill, C., Schopf, P., and Vikhliaev, Y.: MAPL: a high-level programming paradigm to support more rapid and robust encoding of hierarchical trees of interacting high-performance components, in: Proceedings of the 2007 symposium on Component and framework technology in high-performance and scientific computing, ACM, 11–20,
850 https://doi.org/10.1145/1297385.1297388, 2007.

Theurich, G., DeLuca, C., Campbell, T., Liu, F., Saint, K., Vertenstein, M., Chen, J., Oehmke, R., Doyle, J., Whitcomb, T., Wallcraft, A., Iredell, M., Black, T., da Silva, A. M., Clune, T., Ferraro, R., Li, P., Kelley, M., Aleinov, I., Balaji, V., Zadeh, N., Jacob, R., Kirtman, B., Giraldo, F., McCarren, D., Sandgathe, S., Peckham, S., and Dunlap, R.: The Earth System Prediction Suite: Toward a coordinated U.S. modeling capability, Bull. Amer. Meteorol. Soc., 97(7), 1229-1247,
855 https://doi.org/10.1175/BAMS-D-14-00164.1, 2016.

Tzompa-Sosa, Z. A., Mahieu, E., Franco, B., Keller, C. A., Turner, A. J., Helmig, D., Fried, A., Richter, D., Weibring, P., Walega, J., Yacovitch, T. I., Herndon, S. C., Blake, D. R., Hase, F., Hannigan, J. W., Conway, S., Strong, K., Schneider, M., and Fischer, E. V.: Revisiting global fossil fuel and biofuel emissions of ethane, J. Geophys. Res. Atmos., 122(4), 2493-2512, https://doi.org/10.1002/2016JD025767, 2017.

860 United States Environmental Protection Agency: CMAQ (Version 5.2.1) [Software], Zenodo, https://doi.org/10.5281/zenodo.1212601, 2018.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707–11735, https://doi.org/10.5194/acp-10-11707-2010, 2010.

865 van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., Arneth, A., Forrest, M., Hantson, S., Kehrwald, N. M., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Yue, C., Kaiser, J. W., and van der Werf, G. R.: Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015), Geosci. Model Dev., 10, 3329–3357, https://doi.org/10.5194/gmd-10-3329-2017, 2017.

Vinken, G. C. M., Boersma, K. F., Jacob, D. J., and Meijer, E. W.: Accounting for non-linear chemistry of ship plumes in the GEOS-Chem global chemistry transport model, Atmos. Chem. Phys., 11, 11707–11722, https://doi.org/10.5194/acp-11-11707-2011, 2011.

Wang, C., Corbett, J. J., and Firestone, J.: Improving Spatial Representation of Global Ship Emissions Inventories, Environ. Sci. Technol., 42(1), 193-199, https://doi.org/10.1021/es0700799, 2008.

Weng, H.-J., Lin, J.-T., Martin, R., Millet, D. B., Jaeglé, L., Ridley, D., Keller, C., Li, C., Du, M.-X., and Meng, J.: Global high-resolution emissions of soil NOx, sea salt aerosols, and biogenic volatile organic compounds, Sci. Data, 7, 148, https://doi.org/10.1038/s41597-020-0488-5, 2020.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geosci. Model Dev., 4, 625–641, https://doi.org/10.5194/gmd-4-625-2011, 2011.

Xiao, Y., Logan, J. A., Jacob, D. J., Hudman, R. C., Yantosca, R., and Blake, D. R.: Global budget of ethane and regional constraints on U.S. sources, J. Geophys. Res. Atmos., 113(D21), D21306, https://doi.org/10.1029/2007JD009415, 2008.

Xu, W., Wooster, M. J., Roberts, G., and Freeborn, P.: New GOES imager algorithms for cloud and active fire detection and fire radiative power assessment across North, South and Central America, Remote Sens. Environ., 114(9), 1876-1895, https://doi.org/10.1016/j.rse.2010.03.012, 2010.

Zender, C. S., Bian, H., Newman, D.: Mineral Dust Entrainment and Deposition (DEAD) model: Description and 1990s dust climatology, J. Geophys. Res. Atmos, 108(D14), 4416, https://doi.org/10.1029/2002JD002775, 2003.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., and Zhang, Q.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, Atmos. Chem. Phys., 18, 14095–14111, https://doi.org/10.5194/acp-18-14095-2018, 2018.

Zhou, L., Lin, S., Chen, J., Harris, L. M., Chen, X., & Rees, S. L., Toward Convective-Scale Prediction within the Next Generation Global Prediction System, Bull. Amer. Meteorol. Soc., 100(7), 1225-1243, https://doi.org/10.1175/BAMS-D-17-0246.1, 2019.

Zhuang, J., Jacob, D. J., Lin, H., Lundgren, E. W., Yantosca, R. M., Flo Gaya, J., Sulprizio, M. P. and Eastham, S. D.: Enabling High-Performance Cloud Computing for Earth Science Modeling on Over a Thousand Cores: Application to the GEOS-Chem Atmospheric Chemistry Model, J. Adv. Model Earth Sy., 12, e2020MS002064, https://doi.org/10.1029/2020MS002064, 2020.