

Manuscript prepared for Geosci. Model Dev. Discuss.
with version 4.1 of the L^AT_EX class copernicus_discussions.cls.
Date: 28 April 2014

HEMCO v1.0: A versatile, ESMF-compliant component for calculating emissions in atmospheric models

C. A. Keller¹, M. S. Long¹, R. M. Yantosca¹, A. M. Da Silva², S. Pawson², and D. J. Jacob¹

¹School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

²Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Correspondence to: C. A. Keller
(ckeller@seas.harvard.edu)

Abstract

We describe the Harvard-NASA Emission Component version 1.0 (HEMCO), a stand-alone software component for computing emissions in global atmospheric models. HEMCO determines emissions from different sources, regions and species on a user-defined grid and can

- 5 combine, overlay, and update a set of data inventories and scale factors, as specified by the user through the HEMCO configuration file. New emission inventories at any spatial and temporal resolution are readily added to HEMCO and can be accessed by the user without any pre-processing of the data files or modification of the source code. Emissions that depend on dynamic source types and local environmental variables such as wind speed or surface temperature
- 10 are calculated in separate HEMCO extensions.

HEMCO is fully compliant with the Earth System Modeling Framework (ESMF) environment. It is highly portable and can be deployed in a new model environment with only few adjustments at the top-level interface. So far, we have implemented HEMCO in the NASA GEOS-5 Earth System Model (ESM) and in the GEOS-Chem chemical transport model (CTM).

- 15 By providing a widely applicable framework for specifying constituent emissions, HEMCO is designed to ease sensitivity studies and model comparisons, as well as inverse modeling in which emissions are adjusted iteratively. The HEMCO code, extensions, and the full set of emissions data files used in GEOS-Chem are available at <http://wiki.geos-chem.org/HEMCO>.

1 Introduction

- 20 Accurate representation of emissions is essential in global models of atmospheric composition. Models typically rely on gridded emission inventory data, covering global or regional domains, which are often multiplied with scale factors to adjust for different species and temporal variability (Lamarque et al., 2010). New and updated emission inventories are continuously being developed by research groups and agencies, reflecting both improving knowledge and actual
- 25 changes in emissions. Timely incorporation of this new information into atmospheric models is crucial but can involve laborious programming. Here, we present the Harvard-NASA Emission

Component version 1.0 (HEMCO), a software interface for atmospheric models that automates the implementation of new inventories and allows the construction of user-specified combinations of existing inventories and scale factors on a per region and/or per species basis. HEMCO is compliant with the Earth System Modeling Framework (ESMF, Hill et al. (2004)) software environment and thus can serve as a stand-alone emission component in Earth System Models (ESM).

The general approach to determine emission of a given species in global atmospheric models is through a combination of base emissions and multiplicative scale factors. Base emissions are gridded external data generally constructed using a bottom-up approach based on best estimates of activity rates (e.g. fuel consumption) and emission factors (e.g. emitted mass of species per unit mass of fuel) (Granier et al., 2011). They may also include top-down constraints from atmospheric observations (e.g. Mieville et al., 2010). Scale factors applied to these base values adjust emissions at specific times to account for diurnal, day-of-week, seasonal, or year-to-year variability (van Donkelaar et al., 2006; Wang et al., 2010), or for environmental parameters such as wind or temperature (e.g. Zender et al., 2003; Guenther et al., 2012).

HEMCO is highly customizable as it can use a wide range of base emissions and scale factors. A reference library of datasets is included in HEMCO, and user can supplement these with its own alternatives. These inventories need not be of the same grid dimensions or domain. Using the customizable configuration file, the HEMCO core module selects and assembles the emission arrays for the atmospheric model through combination of the selected base emissions and scale factors. More interactive emission modules that depend on gridded source types (e.g. land use, vegetation type) and/or environmental dependent scale factors (e.g. wind speed, surface temperature) are appended to the HEMCO core module as HEMCO extensions, as explained in Section 2.6.

HEMCO is designed so that it is well suited for use in ESMs, where the atmospheric composition module is coupled to modules describing atmospheric dynamics and other components of the Earth system (e.g. oceans, land, cryosphere). ESMs are interdisciplinary endeavors where stewardship of the code is distributed among several research communities, placing an additional hurdle on timely code updates. We have designed HEMCO so it can serve as an emission

component for ESMs through the ESMF interface, thus allowing for seamless updating of emission inventories and extension modules. HEMCO is currently being incorporated into the Goddard Earth Observing System (GEOS-5) ESM of the NASA Global Modeling and Assimilation Office (GMAO) (Molod et al., 2012; Ott et al., 2010; Randles et al., 2013). The HEMCO code,
5 written in FORTRAN 90, along with its current extensions and library of open-source emission inventory databases is available at <http://wiki.geos-chem.org/HEMCO>.

2 Description of HEMCO

2.1 Overview

Figure 1 illustrates the design of the HEMCO core module. HEMCO acts as a coupler between

- 10 a set of emission data files organized in a data library and the external (atmospheric) model.
Based on the specifications of the user configuration file, HEMCO selects the emission files
to be used, schedules and invokes the corresponding data receiving commands, organizes the
resulting data arrays, and calculates the emission fields for a given species and time upon
request. Species definitions and geographical grid points to be used for the emission calculation
15 are specified during initialization of HEMCO. The grid can be 3-D to allow for emissions at
altitude (e.g. from tall stacks, aircraft, etc.), and all emission fields will be returned on this
grid. Grid definitions are typically determined from the external model, even though any grid is
supported. The species definitions are used to identify the species names provided in the con-
figuration file (Section 2.3), and pass the respective emission fields to the atmospheric model.
20 Each species is defined by its name, molecular weight, and corresponding species name/ID used
in the atmospheric model. Additional conversion factors can be defined, e.g. to convert mass of
emitted species into mass of carbon or to map individual species into compound classes (and
vice versa), e.g. for organic compounds.

HEMCO receives all data through the data interface and all data arrays entering HEMCO are
25 already on the requested grid, i.e. data reading and remapping operations are performed outside
of HEMCO core. This facilitates the coupling of HEMCO to different data reading and regrid-

ding algorithms, as discussed further in Section 2.5.

For each gridbox x on the specified grid, HEMCO computes emissions $e_{x,j}(t)$ for requested species j and at time t . It does so by incorporating the emission inventories and scale factor data files selected and prioritized by the user through the configuration file. The resulting emis-

5 sions $e_{x,j}(t)$ are then passed to the external model (Figure 1). Emission calculation may include a combination of different inventories $n \in [1, p]$ covering different geographic domains (e.g. North America, China) and/or emission sectors (e.g. fossil fuel, open fires). For each selected inventory n , the emission $e_{x,j,n}(t)$ is calculated as multiplication of the base value $b_{x,j,n}(t)$ and $m \in [1, q]$ scale factors $s_{x,m}(t)$, as defined in the configuration file:

$$10 \quad e_{x,j,n}(t) = b_{x,j,n}(t) \times \prod_{m=1}^q s_{x,m}(t) \quad (1)$$

Emissions $e_{x,j,n}(t)$ and $b_{x,j,n}(t)$ are in units of mass per unit area per unit time, and scale factors $s_{x,m}(t)$ are unitless. Scale factors represent (1) temporal emission variations including diurnal, seasonal or inter-annual variability; (2) regional masks that restrict the applicability of the base inventory to a given region; or (3) species specific scale factors, e.g. to split lumped 15 organic compound emissions into individual species. Additional scale factors can be applied to have emissions depend on local environmental variables such as temperature or wind speed. These require specifications or functional dependencies and thus special treatment, as will be discussed in Section 2.6.

The final emissions $e_{x,j}(t)$ are composed through addition and/or overwriting of all p employed 20 inventories $e_{x,j,n}(t)$. To determine how inventories of the same species are added and prioritized, each inventory is assigned a category and hierarchy number in the configuration file. Within the same category, inventories of higher priority overwrite lower priority data, while emissions of different categories are added. This system enables the user to prioritize selected regional inventories for a given sector (e.g. anthropogenic, biofuel, etc.) over a default global 25 inventory for the same sector, while still allowing the global inventory to provide information for other sectors. Note that this assumes comparable sector definitions throughout the selected inventories and sectors. Inventories with inconsistent sector definitions can be distributed across

multiple sectors (or lumped into one particular sector) by assigning user-specified weights to each sector.

2.2 Data library

The HEMCO data library contains the data files of all base emissions and scale factors available

5 to users, who may also choose to extend it by adding their own. Depending on the specifications of the configuration file, only a subset of the library is effectively used for emission calculation. Table 1 lists the global and regional emission inventories currently included in the HEMCO library. They correspond to the standard emission settings currently used in the GEOS-Chem chemical transport model (CTM) (Bey et al., 2001). The choice of these emission inventories
10 have been discussed and validated in many publications (e.g. Fairlie et al., 2010; Millet et al., 2010; van Donkelaar et al., 2006; Xiao et al., 2008; Yevich and Logan, 2003).

All files are in the Network Common Data Form (netCDF) format (<http://www.unidata.ucar.edu/software/netcdf/>) - the most commonly used data format in the climate community - and
15 adhere to the COARDS metadata conventions. New inventories following these conventions can be readily added to the data library. On demand, support for other data formats/conventions can be added with relatively little effort through extension of the HEMCO data interface.

All data listed in Table 1 are on rectilinear (lon-lat) grids and become interpolated onto the simulation grid during model execution. At this stage of development, non-regular (e.g. curvilinear) grids are only supported when running HEMCO within the ESMF-environment (see Section
20 2.5).

2.3 HEMCO configuration file

Users select base inventories and scale factors for their simulation through the HEMCO configuration file. Thus, HEMCO enables the user to incorporate new emissions and alter the composition of model emissions without the need to change any source code. A sample configuration file

25 and the corresponding model emission field is shown in Figure 2 for calculating global anthropogenic emissions of NO_x. Default inter-annual global emissions are from EDGAR (Emissions

Database for Global Atmospheric Research, Janssens-Maenhout et al. (2010)), stored in netCDF data file Edgar.nc. Monthly and diurnal scale factors are taken from files MonthScal.nc and DayScal.nc, respectively. Over Asia and Europe, the global EDGAR emissions are overwritten by the monthly regional inventories of Zhang et al. (2009) and the European Monitoring and Evaluation programme (EMEP, Vestreng et al. (2009)), respectively. EMEP covers the period 1985–2007 while the Zhang et al. (2009) data are available for year 2006. The same diurnal scale factors are applied to the two regional inventories as for the EDGAR emissions, and the inter-annual variability of the EDGAR inventory, provided in YearScal.nc, is adapted to the Asian inventory. Finally, NO_x ship emissions from the International Comprehensive Ocean - Atmosphere Data Set (ICOADS, Wang et al. (2008)), available in file Ship.nc, are used in addition to the above-mentioned inventories.

The first section of the configuration file (denoted base emissions) lists the base inventories (see Figure 2). The first column ('Name') is a descriptive field identification name, followed by data reading information consisting of the (netCDF) data filename ('srcFile'), the data variable name ('srcVar'), as well as available time range and temporal resolution ('srcTime'), as described in more detail in Section 2.5. Column 'Species' denotes the emissions species name used by the external model, which is adopted by HEMCO during initialization. It is used to ensure that the requested model species are correctly identified by HEMCO (emissions will be ignored otherwise). Column 'ScalIDs' lists the identification numbers of all scale factors applied to this base inventory, with multiple scale factors separated by the forward slash sign. The numbers refer to the scale factor numbers specified in column 'ScalID' of the second part of the configuration file, where all scale factors and masks are listed. For example, in the configuration file shown in Figure 2, the EDGAR NO_x inventory (line 5) is linked with scale factors 1 (DAY_NOX) and 2 (MONTH_NOX), defined on lines 12 and 13, respectively.

The last two columns of the base data section give the emissions category ('Cat') and hierarchy ('Hier'). In the example of Figure 2, the EDGAR_NOX field (Cat = 1; Hier = 1) is overwritten by the regional ASIA_NOX and EMEP_NOX data (Cat = 1; Hier = 2). The regional inventories are only applied to the region where they are defined, and EDGAR is used everywhere else. The ship emissions SHIP_NOX are given a different category (Cat = 2) and hence are added to the

NO_x field assembled for emission category 1.

The effect of adding the SHIP_NOX emission field to the configuration file is illustrated in the two maps shown in Figure 2. Panel (a) depicts the NOx emission map obtained from running HEMCO without the ship inventory. Panel (b) shows the resulting emission map when line 8 is included.

Sections 2 and 3 of the configuration file list scale factors and region masks. All scale factors and masks are listed with the scale factor identification number ('ScalID'), descriptive field name ('Name') and file attributes ('srcFile', 'srcVar', 'srcTime'). The unitless scale factors are either gridded data obtained from a data file (e.g. geographical variations in diurnal emissions)

- or a spatially uniform scalar directly defined in the configuration file in lieu of a data file. The latter makes it easy to uniformly scale emissions and/or to fractionate lumped emission inventories into different sectors or individual species (e.g. for organic compounds). Masks are binary scale factors (1 inside the region, 0 outside).

15 2.4 Core module and emissions calculations

The operation of HEMCO can be divided into three stages, all invoked by the external model (Figure 1): Initialize, Run, Finalize. Initialize and Finalize are only executed once, at the beginning and end of the model simulation, respectively. The Run command is repeated at every emission time step.

- 20 The core of HEMCO consists of the internal data structures 'FileList' and 'EmisList'. FileList contains the file information of all used base emissions and scale factors, such as data filename, variable, update frequency, etc. It is created in the first stage of HEMCO (Initialize) based on the content of the HEMCO configuration file. In addition to setting FileList, the initialization routine also receives emission species definitions (e.g. species name, molecular weight)
- 25 and specifies the (emission) grid points to be covered by this Central Processing Unit (CPU). In the case of a distributed computing environment, the emissions grid will be broken up across all available CPUs on the system. The grid defined during initialization is preserved over the whole course of the simulation and all emissions are returned on it.

EmisList organizes the 3-D arrays of all base emissions and scale factors, which are stored in individual data structures ('containers') along with information on how these arrays are connected to each other. Each data array covers the specified emission grid and contains the values for the current simulation date. HEMCO automatically updates all arrays as the simulation date

5 advances, based on the update frequency defined in the configuration file.

The second stage (Run) of HEMCO consists of two steps, namely receiving/updating content of the emissions list, and calculating the emissions $e_{x,j}(t)$. The 'receive' command generates a data request, based on the information in FileList, which is sent to the data interface. The returned base emissions and scale factors (which is, $b_{x,j,n}(t)$ and $s_{x,m}(t)$) are on the specified

10 emission grid and unit and become stored in a corresponding data container in EmisList.

In the second step of Run ('calculate'), emissions are directly calculated from the 3-D arrays stored in EmisList according to Eq. 1. For each base inventory n of compound j , emissions $e_{x,j,n}(t)$ are calculated first for every grid point x before all these values are merged into the final emissions $e_{x,j}(t)$, based upon the emission categories and hierarchies given to each inventory.

2.5 Data interface

The data interface provides the link between HEMCO and the input data files. Depending on the employed model environment, this step includes different operations and levels of complexity. When run within ESMF, file reading and data interpolation are performed using the MAPL

20 (Modeling Analysis and Prediction Program Layer) software toolkit built on top of ESMF (<https://modelingguru.nasa.gov/docs/DOC-1118>). In this case, the role of the data interface is to ensure that all files required by HEMCO are correctly identified and registered through MAPL, as well as to connect the final processed data to the HEMCO core module on every emission time step. The ESMF regridding utility supports a wide variety of grids, including rectilinear,

25 curvilinear, and unstructured grids. More details on the HEMCO implementation within a MAPL/ESMF environment is given in Section 3.

If running HEMCO outside of ESMF, e.g. as part of a stand-alone CTM like GEOS-Chem, the data interface needs to perform data reading and remapping operations explicitly. In this case,

- a package of generic subroutines is called through the data interface module. All data reading parameters used by these routines, such as filename, data variable name, update frequency, etc., are specified by the user in the configuration file and become stored in `FileList`. On every emission time step, HEMCO determines the files to be updated - based on the date at the
- 5 current and previous time step and the specified update frequencies - and invokes the reading and remapping routines accordingly. The (netCDF) filename, data variable and time stamp to be read are extracted by HEMCO from columns '`srcFile`', '`srcVar`' and '`srcTime`', respectively, of the configuration file. The time stamp provided in '`srcTime`' has format year/month/day/hour and indicates the available time range as well as the temporal resolution of the configuration file.
- 10 Both discrete dates for time-independent data (e.g. ship emissions in Figure 2: 2000/1/1/0) and time ranges for temporally changing inventories (e.g. EDGAR: 1980-2010/1/1/0) are accepted. Time-uniform data is only read once and the same array is then used for all simulation dates. For time-varying data, the time slice most representative for the current simulation date is used. For example, HEMCO automatically updates EDGAR NO_x data whenever the simulation year
- 15 changes within simulation years 1980 to 2010. Outside of this range, the closest available time slice is used.

At this stage of development, the HEMCO generic reading and remapping routines focus on rectilinear (lon-lat) grids - which is the most commonly used grid type in (global) emission inventories - and all input data is expected to be on such a grid. The fact that all data reading and

20 remapping routines are kept separate from the rest of the HEMCO code (see Figure 1) simplifies connecting HEMCO to other data reading and remapping routines and/or extend existing functionalities, e.g. to use input data in formats different than netCDF or support additional regridding methods.

2.6 Extensions for environmental dependent emissions

- 25 Emission inventories sometimes include dynamic source types and non-linear scale factors that have functional dependencies on local environmental variables, which are best calculated online during execution of the model. Examples are wind dependence of dust emissions (Zender et al., 2003), or the temperature and light dependence of biogenic VOC emissions (Guenther

et al., 2012). In such cases, HEMCO can host environmental independent data sets (source functions) in its data library, but all other scale factors cannot be determined within the HEMCO core module. Instead, users can select a suite of additional modules (extensions) that perform on-line emission calculations based on environmental variables imported from other parts of the
5 ESM.

These extensions take advantage of many of the functionalities of HEMCO. Like the emission data used by the core module, gridded source function data (such as base emissions and environmental independent scale factors) are specified in the configuration file and subsequently become read, organized and stored through the HEMCO FileList and EmisList (Figure 1),
10 except that they are not used for the HEMCO core emission calculation. Instead, these data arrays are requested directly by the respective extension modules and used therein to calculate the emissions for the given process, together with the extension-specific parameterizations. All environmental variables used by the extensions, such as wind speed and surface temperature, are passed to HEMCO from the atmospheric model through the argument interface and then automatically made available to the extension modules.
15

The full set of extensions currently available in HEMCO is given in Table 1. Some of these extensions make model-specific assumptions, e.g. on the dust size bins or VOC speciation, and thus may need some modifications when used in other model environments. User can also write their own extensions and/or import existing model code as additional extensions. A template to
20 facilitate the implementation of new extensions is provided with HEMCO.

Figure 3 illustrates the functioning of the HEMCO extensions for an application that combines three standard emission inventories (Emis001.nc - Emis003.nc) and two extensions (dust emissions and biogenic VOC emissions). The extensions use externally specified, gridded data provided in the configuration file (Erosion.nc, MassFrc.nc, SrcFnc.nc), and compute additional
25 scale factors dependent on environmental variables (wind, temperature, etc.) imported from the atmospheric model.

When calling HEMCO from the external atmospheric model, the HEMCO core code is executed first following the procedure depicted in Figure 1. All six netCDF files listed in the configuration file are received and stored within HEMCO, but only files Emis001-003.nc are used

in HEMCO core for emission calculation. The emissions array calculated in HEMCO core is then passed to the dust emission extension, where dust emissions are calculated as described in Fairlie et al. (2010). The time-invariant (gridded) data required by this extension, namely erodibility and mass fractions per surface type, are obtained from files Erosion.nc and MassFrc.nc,
5 respectively, through the HEMCO core routines. Current meteorological informations like wind speed (u, v) and surface temperature (T) are obtained from the atmospheric model through the argument interface (Figure 3). The calculated dust emissions become added to emissions $e_{x,j}(t)$ previously calculated in HEMCO core, and the combined emission array is then passed to the biogenic VOCs emission extension module. This module requires the gridded source functions
10 (base emissions) for each species - provided by file SrcFnc.nc - along with a suite of meteorological variables (temperature, radiation, etc.) (Guenther et al., 2012) provided by the atmospheric model. The final emission array comprising of the sum of all three emissions arrays is then returned to the atmospheric model.

3 Implementations

15 HEMCO is a stand-alone emissions component that can be readily included into a new model environment. All required adjustments can be done at the interface level between HEMCO and the external model. So far, we have implemented HEMCO in the GEOS-Chem CTM driven by assimilated meteorological data (Bey et al., 2001), and the NASA GEOS-5 Earth System Model from the NASA Goddard Earth Observing System (GEOS-5). The GEOS-Chem implementation
20 uses the ensemble of emission inventories listed in Table 1. New emission inventories are now added to GEOS-Chem through HEMCO, which greatly facilitates model updates. Implementation of HEMCO into the GEOS-5 ESM is done through the ESMF interface. ESMF is a widely used modular software framework for ESMs (Hill et al., 2004). It enables the construction of ESMs by assembly of a number of stand-alone components connected to each
25 other through the ESMF superstructure layer. Components are classified as *gridded* components, which are executed on a discrete grid, and *coulper* components, which connect gridded components and perform input/output operations. Components receive data through a special

superstructure object class (*import state*), and make data available to other components by returning data as object class *export state*.

HEMCO contains all wrapper routines needed to embed it as a gridded component into an ESMF model application. Specifically, all files listed in the HEMCO configuration file are reg-

- 5 istered for data reading at the beginning of a model run. These files then become automatically read and interpolated in space and time through ESMF-generic routines, and HEMCO subsequently imports these arrays through the *import state* object during step ‘Receive’ of the Run stage (Figure 1). Likewise, the *import state* object is used to obtain data from other ESM components needed by some of the HEMCO extensions, e.g. meteorological fields such as wind
- 10 speed and temperature or source type classifications, e.g. vegetation type. All emission arrays calculated within HEMCO are returned as *export state* object so that they are available to other model components (i.e. the transport or chemistry component).

4 Conclusions

HEMCO provides a flexible tool for atmospheric models to compute emissions for different

- 15 sources, regions and species through automatic combination, overlaying, and updating of user-selected inventories and scale factors. New data sets on any spatial grid and temporal resolution can be readily added to HEMCO without modification of the source code. Emissions and scale factors that depend on local environmental parameter such as wind speed and temperature are included through HEMCO extensions.

- 20 A particular advantage of HEMCO is that emissions become regridded and converted to desired units during model execution, which allows a straightforward implementation of new emission inventories with no need to pre-process these data. Thus, HEMCO is well suited for model inter-comparison and emission sensitivity studies. These tasks require running the model with emission data that differ from the default emission settings, which is easily achieved in HEMCO by simply modifying the configuration file.

The strictly modular structure of HEMCO also makes it attractive for inverse modeling applications in which emissions are adjusted iteratively to provide an optimal fit to geospatial obser-

vations (see Enting, 2005). The adjustment factors can be easily implemented into HEMCO as additional scale factors, which are then applied to the base emissions.

HEMCO is ESMF-compliant and can therefore be readily used to compute emissions in Earth System Models (ESM) that rely on the ESMF structure. In such applications, HEMCO makes

- 5 use of the MAPL/ESMF software toolkits to read and interpolate data fields from files as well as to connect HEMCO with other ESM components. HEMCO presently serves as emission component for the GEOS-Chem CTM and for the NASA GEOS-5 ESM (via ESMF). It can be used easily in any Earth system model.

5 Code availability

- 10 The HEMCO code (in FORTRAN 90), current extensions and emission data bases (listed in Table 1), and sample configuration files are available at <http://wiki.geos-chem.org/HEMCO>.

Acknowledgements. This work was supported by the NASA Modeling, Analysis, and Prediction (MAP) Program. The authors would like to thank J.E. Nielsen for his technical support.

References

- 15 Benkovitz, C. M., Scholtz, M. T., Pacyna, J., Tarrasón, L., Dignon, J., Voldner, E. C., Spiro, P. A., Logan, J. A., and Graedel, T. E.: Global gridded inventories of anthropogenic emissions of sulfur and nitrogen, *J. Geophys. Res.*, 101, 29 239–29 253, doi:10.1029/96JD00126, 1996.
- 20 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.*, 106, 23 073–23 095, doi:10.1029/2001JD000807, 2001.
- Bond, T. C., Bhardwaj, E., Dong, R., Jogani, R., Jung, S., Roden, C., Streets, D. G., and Trautmann, N. M.: Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850–2000, *Global Biogeochem. Cy.*, 21, GB2018, doi:10.1029/2006GB002840, 2007.
- 25 Diehl, T., Heil, A., Chin, M., Pan, X., Streets, D., Schultz, M., and Kinne, S.: Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and SO₂ from 1980 to 2010 for

hindcast model experiments, *Atmos. Chem. Phys. Discuss.*, 12, 24 895–24 954, doi:10.5194/acpd-12-24895-2012, 2012.

Enting, I.: Inverse Problems in Atmospheric Constituent Transport, Cambridge University Press, New York, 2005.

- 5 Eyring, V., Köhler, H. W., van Aardenne, J., and Lauer, A.: Emissions from international shipping: 1. The last 50 years, *J. Geophys. Res.*, 110, D17305, doi: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, *Atmospheric Chemistry and Physics*, 10, 3999–4012, doi:10.5194/acp-10-3999-2010, 2010.
- 10 Gong, S. L.: A parameterization of sea-salt aerosol source function for sub- and super-micron particles, *Global Biogeochem. Cy.*, 17, 1097, doi:10.1029/2003GB002079, 2003.
- 15 Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., Denier van der Gon, H., Frost, G., Heil, A., Kaiser, J., Kinne, S., Klimont, Z., Kloster, S., Lamarque, J.-F., Lioussse, C., Masui, T., Meleux, F., Mieville, A., Ohara, T., Raut, J.-C., Riahi, K., Schultz, M., Smith, S., Thompson, A., Aardenne, J., Werf, G., and Vuuren, D.: Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period, *Climatic Change*, 109, 163–190, doi:10.1007/s10584-011-0154-1, 2011.
- 20 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, doi:10.5194/gmd-5-1471-2012, 2012.
- 25 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, doi:10.1109/MCISE.2004.1255817, 2004.
- 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.*, 12, 7779–7795, doi:10.5194/acp-12-7779-2012, 2012.
- 30 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, doi:10.5194/acp-11-3137-2011, 2011.
- Janssens-Maenhout, A., Petrescu, A., Muntean, M., and Blujdea, V.: Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements, The National Academies Press, Washington, DC., 2010.

- Johnson, M. T.: A numerical scheme to calculate temperature and salinity dependent air-water transfer velocities for any gas, *Ocean Sci.*, 6, 913–932, doi:10.5194/os-6-913-2010, 2010.
- Kuhns, H., Knipping, E. M., and Vukovich, J.: Development of a United States-Mexico Emissions Inventory for the Big Bend Regional Aerosol and Visibility Observational (BRAVO) Study, *J. Air Waste Ma.*, 55, 677–692, doi:10.1080/10473289.2005.10464648, 2005.
- Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Lioussse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, *Atmos. Chem. Phys.*, 10, 7017–7039, doi:10.5194/acp-10-7017-2010, 2010.
- 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 Br_y: a global modeling study of CHBr₃ and CH₂Br₂, *Atmos. Chem. Phys.*, 10, 2269–2286, doi:10.5194/acp-10-2269-2010, 2010.
- Mieville, A., Granier, C., Lioussse, C., Guillaume, B., Mouillet, F., Lamarque, J.-F., Grégoire, J.-M., and Pétron, G.: Emissions of gases and particles from biomass burning during the 20th century using satellite data and an historical reconstruction, *Atmos. Environ.*, 44, 1469 – 1477, doi:10.1016/j.atmosenv.2010.01.011, 2010.
- Millet, D. B., Guenther, A., Siegel, D. A., Nelson, N. B., Singh, H. B., de Gouw, J. A., Warneke, C., Williams, J., Eerdeken, 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, *Atmospheric Chemistry and Physics*, 10, 3405–3425, doi:10.5194/acp-10-3405-2010, 2010.
- Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I., and Eichmann, A.: The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna, Tech. Rep. NASA/TM-2012-104606-VOL-28; GSFC.TM.01153.2012, NASA Goddard Space Flight Center; Greenbelt, MD, United States, 2012.
- 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, doi:10.1029/2012JD017934, 2012.
- Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., Boutin, J., and Upstill-Goddard, R. C.: In situ evaluation of air-sea gas exchange parameterizations using novel con-

servative and volatile tracers, *Global Biogeochem. Cy.*, 14, 373–387, doi:10.1029/1999GB900091, 2000.

Ott, L. E., Duncan, S., Pawson, P., Colarco, M., Chin, C., Randles, C., Diehl, T., and Nielsen, J. E.: Influence of the 2006 Indonesian biomass burning aerosols on tropical dynamics studied with the
5 GEOS-5 AGCM, *J. Geophys. Res.*, 115, D14121, doi:10.1029/2009JD013181, 2010.

Randles, C. A., Colarco, P. R., and Silva, A. d.: Direct and semi-direct aerosol effects in the NASA
GEOS-5 AGCM: aerosol-climate interactions due to prognostic versus prescribed aerosols, *J. Geophys. Res.*, 118, doi:10.1029/2012JD018388, 149–169, 2013.

Schultz, M. G., Heil, A., Hoelzemann, J. J., Spessa, A., Thonicke, K., Goldammer, J. G., Held, A. C.,
10 Pereira, J. M. C., and van het Bolscher, M.: Global wildland fire emissions from 1960 to 2000, *Global Biogeochem. Cy.*, 22, GB2002, doi:10.1029/2007GB003031, 2008.

Stettler, M., Eastham, S., and Barrett, S.: Air quality and public health impacts of {UK} airports. Part I: Emissions, *Atmos. Environ.*, 45, 5415 – 5424, doi:10.1016/j.atmosenv.2011.07.012, 2011.

Streets, D. G., Bond, T. C., Carmichael, G. R., Fernandes, S. D., Fu, Q., He, D., Klimont, Z., Nelson,
15 S. M., Tsai, N. Y., Wang, M. Q., Woo, J.-H., and Yarber, K. F.: An inventory of gaseous and primary aerosol emissions in Asia in the year 2000, *J. Geophys. Res.*, 108, 8809, doi:10.1029/2002JD003093, 2003.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton,
20 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, 11 707–11 735, doi:10.5194/acp-10-11707-2010, 2010.

van Donkelaar, A., Martin, R. V., and Park, R. J.: Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing, *J. Geophys. Res.*, 111, D21201, doi:10.1029/2005JD006996, 2006.

25 Vestreng, V., Ntziachristos, L., Semb, A., Reis, S., Isaksen, I. S. A., and Tarrasón, L.: Evolution of NO_x emissions in Europe with focus on road transport control measures, *Atmos. Chem. Phys.*, 9, 1503–1520, doi:10.5194/acp-9-1503-2009, 2009.

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, 11 707–11 722, doi:10.5194/acp-11-11707-2011, 2011.

30 Wang, C., Corbett, J. J., and Firestone, J.: Improving Spatial Representation of Global Ship Emissions Inventories, *Environ. Sci. Technol.*, 42, 193–199, doi:10.1021/es0700799, 2008.

- Wang, X., Liang, X.-Z., Jiang, W., Tao, Z., Wang, J. X., Liu, H., Han, Z., Liu, S., Zhang, Y., Grell, G. A., and Peckham, S. E.: WRF-Chem simulation of East Asian air quality: Sensitivity to temporal and vertical emissions distributions, *Atmos. Environ.*, 44, 660 – 669, doi:10.1016/j.atmosenv.2009.11.011, 2010.
- 5 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.*, 113, D21306, doi:10.1029/2007JD009415, 2008.
- Yevich, R. and Logan, J. A.: An assessment of biofuel use and burning of agricultural waste in the developing world, *Global Biogeochem. Cy.*, 17, 1095, doi:10.1029/2002GB001952, 2003.
- 10 Zender, C. S., Bian, H., and Newman, D.: Mineral Dust Entrainment and Deposition (DEAD) model: Description and 1990s dust climatology, *J. Geophys. Res.*, 108, 4416, doi:10.1029/2002JD002775, 2003.
- Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006
15 for the NASA INTEX-B mission, *Atmos. Chem. Phys.*, 9, 5131–5153, doi:10.5194/acp-9-5131-2009, 2009.

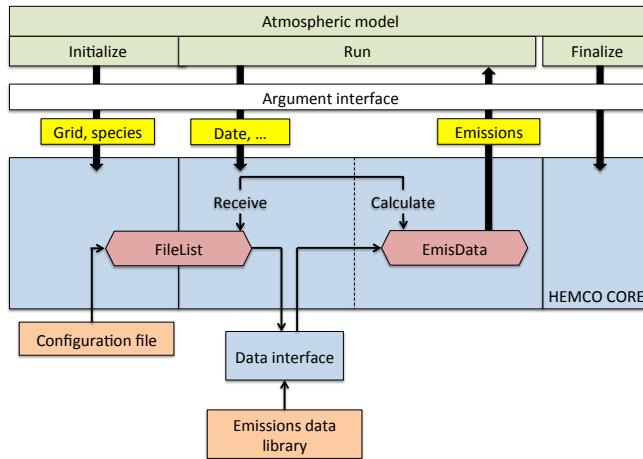


Fig. 1. Overview of the HEMCO core module, which acts as coupler between the atmospheric model and emission files organized in a data library. All emission calculations are based on the information provided in the HEMCO configuration file, which is read and stored in the internal file list ('FileList') during initialization (`Initialize`). By modifying the content of the data library and configuration file, users can readily add new emission data and change emission calculation settings. HEMCO calculates emissions for a given species, date, and grid in two steps: first, all emission data to be updated are identified and selected from 'FileList' and received through the data interface ('Receive'). The returned data become stored in the 'EmisData' list, from which the final emission array is assembled in the second step of RUN ('Calculate'). This array is then returned to the atmospheric model. Routine `FINALIZE`, called at the end of a model run, cleanly removes all internal data.

```

1  ### BEGINNING OF HEMCO CONFIGURATION FILE ####
2
3  ### Base emissions ####
4  Name      srcFile      srcVar      srcTime      Species      ScalIDs  Cat  Hier
5  EDGAR_NOX Edgar.nc    NOX         1980-2010/1/1/0 NOx          1/2      1     1
6  ASIA_NOX   Asia.nc    NOX         2006/1-12/1/0  NOx          1/3/91   1     2
7  EMEP_NOX   Emep.nc    NOX         1985-2007/1-12/1/0 NOx          1/92     1     2
8  SHIP_NOX   Ship.nc    NOX         2000/1/0/0      NOx          -        2     1
9
10 ### Scale factors ####
11 ScalID      Name      srcFile      srcVar      srcTime
12 1           DAY_NOX  DayScal.nc  NOxScale    2000/1/1/1-24
13 2           MONTH_NOX MonthScal.nc NOxScale    2000/1-12/1/0
14 3           YEAR_NOX  YearScal.nc NOxScale    1980-2000/1/1/0
15
16 ### Masks ####
17 ScalID      Name      srcFile      srcVar      srcTime
18 91          ASIA_MASK AsiaMask.nc mask        2000/1/1/0
19 92          EMEP_MASK EmepMask.nc mask        2000/1/1/0
20
21 ### END OF HEMCO CONFIGURATION FILE ####

```

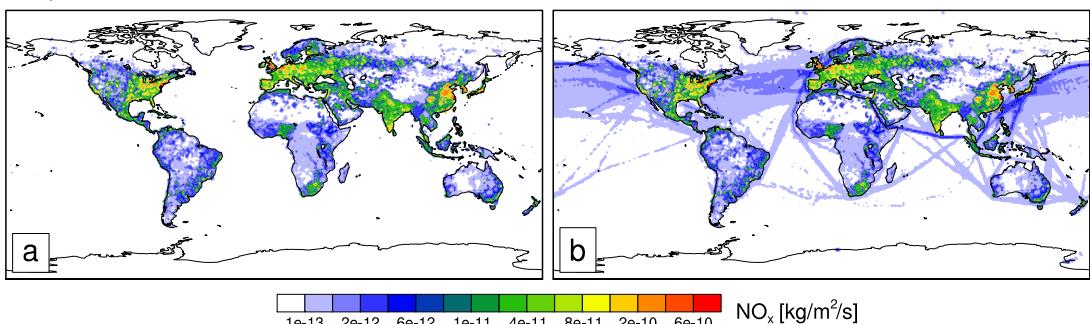


Fig. 2. Sample HEMCO configuration file and resulting emission fields calculated by HEMCO. Emission inventories (base emissions) are listed in lines 5-8 with the identification name ('Name'), (netCDF) filename ('srcFile'), file variable of interest ('srcVar'), and temporal resolution ('srcTime': year/month/day/hour). The atmospheric model species name is given in the 5th column ('Species'), and all scale factors to be linked to the base emissions are listed in the 6th column ('ScalIDs'), separated by a forward slash. The emission category and hierarchy are defined in columns 'Cat' and 'Hier', respectively. Scale factors and region masks are listed in lines 12-19, starting with the scale factor ID ('ScalID', corresponding to the IDs given in the base emission section), the identification name ('Name'), and the data file information ('srcFile', 'srcVar', 'srcTime', as for base emissions). Panel (a) and (b) show the emission maps calculated by HEMCO without and with the ship NO_x inventory (in bold, line 8), respectively. See Section 2.3 for more details.

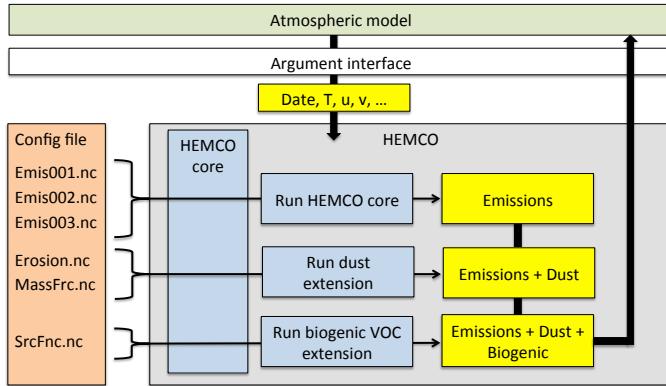


Fig. 3. Use of HEMCO extensions to compute emissions with functional dependencies on local environmental variables. Three standard emission inventories (Emis001.nc - Emis003.nc) and two extensions (dust emissions and biogenic VOC emissions) are combined. The extensions use external gridded data (Erosion.nc, MassFrc.nc, SrcFnc.nc) and compute additional scale factors dependent on temperature (T), wind (u, v), etc., imported from the atmospheric model through the argument interface. All selected HEMCO extension are sequentially called after the HEMCO core module, and the emission arrays calculated therein become added to the emissions previously calculated in HEMCO core. See Section 2.6 for more details.

Table 1. Sample of emission inventories currently in the HEMCO library. The horizontal resolution of the extensions is determined by the meteorological input data.

Coverage	Species ^a	Resolution (lat × lon)	Reference
Global	CO, NO _x , SO ₂	0.1° × 0.1°	EDGAR (Janssens-Maenhout et al., 2010)
Global	VOCs	0.5° × 0.5°	RETRO (Schultz et al., 2008)
Global	NH ₃	1° × 1°	GEIA (Benkovitz et al., 1996)
Global	OC, BC	1° × 1°	Bond et al. (2007)
Global	C ₂ H ₆	2° × 2.5°	Xiao et al. (2008)
Global	CHBr ₃ , CH ₂ Br ₂	2° × 2.5°	Liang et al. (2010)
Global	Biofuel CO, NO _x , VOCs	2° × 2.5°	Yevich and Logan (2003)
Global	Volcanic SO ₂	0.5° × 0.5°	Diehl et al. (2012)
Global	Fire NO _x , CO, SO ₂ , VOCs, BC/OC, NH ₃	0.5° × 0.5°	GFED-3 (van der Werf et al., 2010)
Global	Ship SO ₂	1° × 1°	ARCTAS (Eyring et al., 2005)
Global	Ship CO, NO _x , SO ₂	1° × 1°	ICOADS (Wang et al., 2008)
Global	Aircraft NO _x , CO, SO ₂ , VOCs, BC, OC	1° × 1°	AEIC (Stettler et al., 2011)
US	NO _x , CO, SO ₂ , VOCs, NH ₃	0.25° × 0.3125°	EPA (http://www.epa.gov/ttn/chief/)
Canada	NO _x , CO, SO ₂ , NH ₃	1° × 1°	CAC (http://www.ec.gc.ca/inrp-npri/)
Mexico	NO _x , CO, SO ₂	1° × 1°	BRAVO (Kuhns et al., 2005)
Europe	NO _x , CO, SO ₂ , VOCs, NH ₃	0.5° × 0.5°	EMEP (Vestreng et al., 2009)
East Asia	NO _x , CO, SO ₂ , VOCs, NH ₃	0.5° × 0.5°	Zhang et al. (2009); Streets et al. (2003)
Extensions			
Global	Biogenic VOCs		MEGAN (Guenther et al., 2012)
Global	Mineral dust aerosols ^b		Zender et al. (2003); Fairlie et al. (2010)
Global	Accumulation and coarse mode sea salt aerosols ^c		Jaeglé et al. (2011); Gong (2003)
Global	Oceanic DMS, C ₃ H ₆ O		Johnson (2010); Nightingale et al. (2000)
Global	Ship NO _x , O ₃ , HNO ₃		PARANOX (Vinken et al., 2011)
Global	Lightning NO _x		Murray et al. (2012)
Global	Soil and fertilizer NO _x		Hudman et al. (2012)

^aVOCs = volatile organic compounds; OC = organic carbon aerosol; BC = black carbon aerosol.

^b4 size bins with radii 0.1 – 1.0, 1.0 – 1.8, 1.8 – 3.0, and 3.0 – 6.0 μm, respectively (Fairlie et al., 2010).

^cEffective radii are 0.01 – 0.5 μm and 0.5 – 8 μm for accumulation and coarse mode, respectively (Jaeglé et al., 2011).