



Development of a grid-independent GEOS-chem chemical transport model as an atmospheric chemistry module for Earth System Models

Citation

Long, M. S., R. Yantosca, J. E. Nielsen, C. A. Keller, A. da Silva, M. P. Sulprizio, S. Pawson, and D. J. Jacob. 2014. "Development of a Grid-Independent GEOS-Chem Chemical Transport Model as an Atmospheric Chemistry Module for Earth System Models." *Geoscientific Model Development Discussions* 7 (6): 7505–7524. doi:10.5194/gmdd-7-7505-2014.

Published version

<https://doi.org/10.5194/gmdd-7-7505-2014>

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:14004548>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles (OAP), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Development of a Grid-Independent GEOS-Chem Chemical Transport Model as an atmospheric chemistry module for Earth System Models.

M.S. Long, R. Yantosca, J. E. Nielsen, C.A. Keller, A. da Silva, M.P. Sulprizio., S. Pawson, D. J. Jacob

Abstract

The GEOS-Chem global chemical transport model (CTM), used by a large atmospheric chemistry research community, has been re-engineered to also serve as an atmospheric chemistry module for Earth System Models (ESMs). This was done using an Earth System Modeling Framework (ESMF) interface that operates independently of the GEOS-Chem scientific code, permitting the exact same GEOS-Chem code to be used as an ESM module or as a stand-alone CTM. In this manner, the continual stream of updates contributed by the CTM user community is automatically passed on to the ESM module, which remains state-of-science and referenced to the latest version of the standard GEOS-Chem CTM. A major step in this re-engineering was to make GEOS-Chem grid-independent, i.e., capable of using any geophysical grid specified at run time. GEOS-Chem data “sockets” were also created for communication between modules and with external ESM code via the ESMF. The grid-independent, ESMF-compatible GEOS-Chem is now the standard version of the GEOS-Chem CTM. It has been implemented as an atmospheric chemistry module into the NASA GEOS-5 ESM. The coupled GEOS-5/GEOS-Chem system was tested for scalability and performance with a tropospheric oxidant-aerosol simulation (120 coupled species, 66 transported tracers) using 48-240 cores and MPI parallelization. Numerical experiments demonstrate that the GEOS-Chem chemistry module scales efficiently for the number of processors tested. Although inclusion of atmospheric chemistry in ESMs is computationally expensive, the excellent scalability of the chemistry module means that the relative cost goes down with increasing number of MPI processes.

1. Introduction

Global modeling of atmospheric chemistry involves solution of the 3-D continuity equations for the concentrations of chemical species including the effects of emissions, transport, chemistry, and deposition. This is commonly done with Chemical Transport Models (CTMs) driven by input meteorological data and surface boundary conditions. CTMs are relatively simple computational tools because the chemical continuity equations are solved without coupling to atmospheric dynamics. They are adequate for many applications and play a central role in advancing knowledge of atmospheric chemistry. However, there is also increasing demand for atmospheric chemistry to be implemented as a coupled module in Earth System Models (ESMs) that represent the ensemble of processes affecting the Earth system. Here we describe a software framework through which the state-of-science GEOS-Chem CTM can be implemented seamlessly as a module in ESMs, so that the stand-alone CTM and the ESM module use exactly the same code. We describe the deployment of this capability in the NASA Goddard Earth Observing System (GEOS) developed at NASA’s Global Modeling and Assimilation Office (GMAO).

GEOS-Chem (<http://www.geos-chem.org>) is a shared-memory parallel (OpenMP) global 3-D Eulerian CTM driven by assimilated meteorological data (Bey et al., 2001). It is used by over 100 research groups worldwide for a wide range of applications including simulation of tropospheric oxidants ((Mao et al., 2013), aerosols (Fairlie et al., 2007; Jaeglé et al., 2011; Park et al., 2004), carbon gases (Nassar et al., 2010; Wang et al., 2004; Wecht et al., 2014), mercury (Holmes et al., 2010; Selin et al., 2008), and stratospheric chemistry (Eastham et al., 2014; Murray et al., 2012). Development of GEOS-Chem is

45 based on core principles of open-source code development, modular structure, nimble approach to
 46 innovation, strong version control and benchmarking, extensive documentation, and user support . The
 47 large user base permits extensive model diagnosis and generates a continual stream of new developments
 48 to maintain the model at the forefront of the science. Implementation of these developments in the
 49 standard GEOS-Chem code can be done quickly and efficiently because of the simplicity of the code and
 50 the common interests of the user community. Maintaining state-of-science capability is more challenging
 51 in ESMs because of complexity of managing the central code and the need for dialogue across research
 52 communities to prioritize model development. On the other hand, CTMs such as GEOS-Chem have more
 53 difficulty staying abreast of high-performance computing (HPC) technology because of limited software
 54 engineering resources.

55 Here we present a re-engineered standard version of the GEOS-Chem CTM capable of serving as a
 56 flexible atmospheric chemistry module for ESMs. A key innovation is that GEOS-Chem is now grid-
 57 independent, i.e., it can be used with any geophysical grid. The same standard GEOS-Chem code can be
 58 integrated into ESMs through the Earth System Modeling Framework (ESMF, Hill et al., 2004) interface,
 59 or used as before as a stand-alone CTM driven by assimilated meteorological data. The re-engineered
 60 grid-independent flexibility has been integrated into the standard open-code version of the GEOS-Chem
 61 CTM. The exact same scientific code in the GEOS-Chem CTM now serves as atmospheric chemistry
 62 module in the GEOS-5 ESM. Scientific updates to the standard GEOS-Chem CTM contributed by its user
 63 community are immediately integrated into the GEOS-5 ESM, so that the ESM effortlessly remains state-
 64 of-science and traceable to the latest standard version of GEOS-Chem.

65

66 2. *Grid-Independent GEOS-Chem Model Description*

67 The GEOS-Chem CTM consists of four modules executing operations for chemistry and dry deposition,
 68 emissions, wet deposition, and transport (Fig. 1). GEOS-Chem solves the general Eulerian form of the
 69 coupled continuity equations for m chemical species with number density vector $\mathbf{n} = (n_1, \dots, n_m)^T$

$$70 \quad \frac{\partial n_i}{\partial t} = -\nabla \cdot (n_i \mathbf{U}) + P_i(\mathbf{n}) - L_i(\mathbf{n}) \quad i \in [1, m] \quad (1)$$

71

72 Here \mathbf{U} is the wind vector (including sub-grid components parameterized as turbulent diffusion and
 73 convection). and $P_i(\mathbf{n})$ and $L_i(\mathbf{n})$ are the local production and loss rates of species i including terms to
 74 describe chemical reactions, aerosol microphysics, emissions, precipitation scavenging, and dry
 75 deposition. In GEOS-Chem, as in all 3-D CTMs, equation (1) is solved by operator splitting to separately
 76 and successively apply concentration updates over finite time steps from a transport operator

77

$$78 \quad \frac{\partial n_i}{\partial t} = -\nabla \cdot (n_i \mathbf{U}) \quad i \in [1, m] \quad (2)$$

79

80 and a local operator (commonly called chemical operator)

81

$$82 \quad \frac{dn_i}{dt} = P_i(\mathbf{n}) - L_i(\mathbf{n}) \quad i \in [1, m] \quad (3)$$

83

84 The transport operator includes no coupling between species, while the chemical operator has no spatial
 85 coupling. The transport operator is further split into 1-D advection operators, a convection operator, and a
 86 boundary layer mixing operator. Operator splitting breaks down the multi-dimensionality of the coupled
 87 system (1) and enables numerical solution by finite differencing. The chemical operator in GEOS-Chem

88 is further split into chemistry and dry deposition, emissions, and wet deposition modules for
89 computational convenience.

90 The transport operators in the standard GEOS-Chem CTM are applied on fixed latitude-longitude grids
91 (e.g. Wu et al. 2007). When integrated into an ESM, GEOS-Chem does not need to calculate its own
92 transport; this is done separately in the ESM as part of the simulation of atmospheric dynamics, where
93 transport of chemical species is done concurrently with transport of meteorological variables. Thus the
94 ESM only uses GEOS-Chem to solve the chemical operator (3) over specified time steps. The GEOS-
95 Chem chemical operator must in turn be able to accommodate any ESM grid and return concentration
96 updates on that grid.

97 The chemical operator has no spatial dimensionality (0-D) and could in principle be solved
98 independently for all grid points of the ESM. However, grouping the grid points by column is more
99 efficient as it permits simultaneous calculation of radiative transfer, precipitation scavenging, and
100 vertically distributed emissions for all grid points within the column. Thus we take a 1-D vertical column
101 as the minimum set of grid points to be handled by a call to the chemical operator. Chemical operator
102 updates for a given column can be completed without information from neighboring columns. Solving for
103 the chemical operator column by column reduces memory overhead and facilitates scalable single
104 program, multiple data (SPMD; Cotronis and Dongarra, 2001) parallelization in a distributed computing
105 environment using the Message Passing Interface (MPI). It may sometimes be preferable to apply the
106 chemical operator to ensembles of columns, grouped independent of geography, to balance the
107 computational burden and achieve performance gains (Long et al., 2013).

108 Prior to this work, the horizontal grid of GEOS-Chem was defined at compile time from a limited
109 selection of fixed latitude-longitude grids ($1/4^\circ \times 5/16^\circ$, $1/2^\circ \times 2/3^\circ$, $2^\circ \times 2.5^\circ$, $4^\circ \times 5^\circ$) compatible with the
110 advection module and offline meteorological fields. Our goal here was to re-engineer the existing GEOS-
111 Chem code to accept any horizontal grid defined at runtime. The horizontal grid would be able to span the
112 entire global domain, represent a single column to be calculated on a single compute node, or represent
113 any collection of columns defined by their location. This permits use of the same scientific code for stand-
114 alone CTM and coupled ESM applications.

115

116 *2.1 Code Modularization & Structure*

117 In order for the GEOS-Chem code to permit run-time horizontal grid definition, much of the
118 FORTRAN-77 code base was updated to leverage Fortran-90 capabilities. This included extensive
119 conversion of static to dynamically-allocatable arrays, and introduction of pointer-based derived data
120 types. Data flow into, through and out of GEOS-Chem's routines was reconfigured to use derived-type
121 objects passed to routines as arguments in place of publicly-declared global-scope variables. This
122 permitted the bundling of data structures with similar functionality into common interfaces (data
123 "sockets") that simplify module communication within GEOS-Chem and coupling to external
124 components through the ESMF interface (see Section 2.2). Three sockets are defined: a meteorology &
125 physics socket, a chemistry socket, and an input options socket. The meteorology & physics socket
126 provides data defining geophysical state variables and arrays. This includes temperature, pressure,
127 humidity, wind fields, and many others. The chemistry socket provides data structures for chemical
128 species including indexing, species names, and concentrations. The input options socket provides runtime
129 information such as calendar, grid dimensions, diagnostic definitions, and locations of offline information
130 stored on disk. Together, these sockets incorporate all of the quantities and fields necessary for coupling
131 to and driving modules within GEOS-Chem.

132 The GEOS-Chem code includes specific hooks to accommodate the ESMF interface and permit
133 coupling with external data streams. These hooks do not interfere with GEOS-Chem’s scientific operation
134 and are used exclusively in grid, I/O, and utility operations. They can remain invisible to the scientific
135 programmer. There are three hooks invoked as C-preprocessor macros: *ESMF_*, *EXTERNAL_GRID*, and
136 *EXTERNAL_FORCING*. Code bounded by these macros is neither compiled nor executed unless the
137 specific macro is enabled at compile time. The *ESMF_* macro bounds code specific for the ESMF. The
138 *EXTERNAL_GRID* macro bounds code that allows GEOS-Chem to operate on an externally defined and
139 initialized grid (e.g. by an ESM). The *EXTERNAL_FORCING* macro bypasses GEOS-Chem’s internal,
140 offline data I/O operations necessary for CTM operation, and replaces them with ESMF-based I/O. Users
141 do not need to have the ESMF installed in order to run GEOS-Chem as a stand-alone CTM. The system
142 reverts to the standard GEOS-Chem CTM code relying on the legacy module interface when compiled
143 without these hooks enabled. It is fully backward-compatible with the current GEOS-Chem CTM
144 operating environment (Fig. 1).

145 The recently developed Harvard-NASA Emissions Component HEMCO ([http://wiki-geos-](http://wiki-geos-chem.org/HEMCO/)
146 [chem.org/HEMCO/](http://wiki-geos-chem.org/HEMCO/)) is used for emission calculations (Keller et al., 2014). HEMCO is a Fortran-90
147 based, ESMF compliant, highly customizable module that uses base emissions and scale factors from a
148 reference database to construct time-dependent emission field arrays. Emission inventories and scale
149 factors are selected by the user in a HEMCO-specific configuration file. Emission inventories for
150 different species and source types need not be of the same grid dimensions or domain. HEMCO was
151 designed by Keller et al. (2014) as a flexible general tool for facilitating the implementation and update of
152 emission inventories in CTMs and ESMs.

153 154 2.2 ESMF Interface

155 GEOS-Chem interfaces with external ESMs using the ESMF. The ESMF is an open-source software
156 application programming interface that provides a standardized high-performance software infrastructure
157 for use in ESM design. It facilitates HPC, portability, and interoperability in Earth science applications
158 (Collins et al., 2005).

159 GEOS-Chem is executed within the ESMF as a gridded component. The gridded component is the basic
160 element of an ESMF-based program, and is defined as a set of discrete scientific and computational
161 functions that operate on a geophysical grid. Likewise, other components of the Earth system are
162 implemented as gridded components (e.g. atmospheric dynamics, ocean dynamics, terrestrial
163 biogeochemistry, etc.).

164 Each gridded component consists of a routine establishing ESMF-specific services, and *Initialize*, *Run*,
165 and *Finalize* operations methods for gridded component execution by the ESMF. The *Initialize* method is
166 executed once at the beginning of the time step and initializes component-specific runtime parameters.
167 The *Run* method interfaces local data structures with ESMF States (see below) and executes the GEOS-
168 Chem code. The *Finalize* method wraps up code execution, closes any remaining open files, finalizes I/O
169 and profiling processes, and flushes local memory.

170 Gridded components exchange information with each other through States. A State is an ESMF derived
171 type that can contain multiple types of gridded and non-gridded information (Collins et al., 2005; Suarez
172 et al., 2013). An ESMF gridded component is associated with an *Import* State and an *Export* State. The
173 *Import* State provides access to data created by other gridded components. The *Export* State contains data
174 that a component generates and makes available to other components. In the ESMF-enabled GEOS-
175 Chem, data are passed into and out of the GEOS-Chem gridded component via interfacing an appropriate

176 State with a corresponding GEOS-Chem data socket, making these data available within GEOS-Chem or
177 to other ESM gridded components (see Section 2.1).

178 The ESMF was implemented within GEOS-Chem as an independent layer that operates on top of the
179 CTM code. It includes code for interfacing with and executing GEOS-Chem as an ESMF gridded
180 component. When coupling GEOS-Chem to an ESM, the GEOS-Chem transport modules are excluded
181 and only those modules necessary to solve Eq. (3) are used. Coupling specifically to the GEOS-5 ESM
182 required an adaptation of GEOS-Chem's ESMF interface for the GMAO's Modeling, Analysis and
183 Prediction Layer (MAPL) extension (Suarez et al., 2013). MAPL is otherwise not required for GEOS-
184 Chem.

185

186 3. Implementation, Performance, and Scalability

187 The ESMF-enabled GEOS-Chem was embedded within the NASA GEOS-5 ESM (version Ganymed-
188 4.0). The GEOS-5 ESM is the forward model of the GEOS-5 atmospheric data assimilation system
189 (GEOS-DAS) (Ott et al., 2009; Rienecker et al., 2008). The system is built upon on an ESMF framework,
190 and uses a combination of distributed memory (MPI) and, in some cases, hybrid distributed/shared
191 memory parallelization. The dynamical core used here is based on Lin (2004), and operates on horizontal
192 grid resolutions ranging from $2^\circ \times 2.5^\circ$ to $0.25^\circ \times 0.3125^\circ$, with 72 vertical layers up to 0.01 hPa. Ocean
193 surface and sea-ice boundaries are prescribed. The land and snow interfaces are based on Koster et al. (
194 2000) and (Stieglitz et al., 2001), respectively. For the coupled simulations, GEOS-5 ESM native
195 dynamics and moist physics (including cloud processes and in-cloud scavenging) are applied to the
196 GEOS-Chem chemical tracers.

197 All coupled GEOS-5/GEOS-Chem simulations were performed on the Discover system at the NASA
198 Goddard Space Flight Center (http://www.nccs.nasa.gov/discover_front.html), using 12-core (dual hex-
199 core) 2.8 GHz Intel Xeon Westmere (X5660) compute nodes equipped with 24 GB RAM, and an
200 Infiniband DDR interconnect using the Intel compiler suite (v. 13.1.1) and MVAPICH2 (v. 1.8.1). GEOS-
201 Chem's shared-memory (OpenMP) parallelization was disabled.

202 The coupled GEOS-5/GEOS-Chem system was tested on $2^\circ \times 2.5^\circ$ and $0.5^\circ \times 0.625^\circ$ grids with a standard
203 oxidant-aerosol simulation using 120 chemical species of which 66 are transported ("chemical tracers").
204 Radical species with very short chemical lifetimes are not transported. The chemistry module used the
205 RODAS-3 (4-stage, order 3(2), stiffly accurate) solver with self-adjusting internal time step (Hairer and
206 Wanner, 1996) as part of the Kinetics Pre-processor (KPP, Eller et al., 2009; Sandu and Sander, 2005).
207 KPP was implemented with its supplied linear algebra (BLAS Level-1) routines in place. The $2^\circ \times 2.5^\circ$
208 simulation used a time step of 1800 seconds for all operations. For the $0.5^\circ \times 0.625^\circ$ simulation, chemistry
209 and system-operation time steps were both 450 seconds. Dynamics, physics, and radiation time steps were
210 900 seconds. For both simulations, the atmosphere used 72 vertical hybrid-sigma (pressure) levels.
211 Simulations were run for 31 days initialized on July 1, 2006. All chemical tracers were initialized from
212 output of a GEOS-Chem CTM (v9-02) simulation.

213 The $2^\circ \times 2.5^\circ$ coupled simulation was used to test scalability of the coupled system and for comparison to
214 the GEOS-Chem CTM. Scalability simulations were run with 48, 96, 144, 192, and 240 total MPI
215 processes operating on 12x4, 12x8, 12x12, 16x12, and 16x15 (lat x lon) contiguous grid point
216 subdomains, respectively. This represents a set of five simulations $i \in [1, 5]$. For comparison, the offline
217 GEOS-Chem CTM (v9-02) was run on 8 shared-memory processes at $2^\circ \times 2.5^\circ$ resolution using 8-core 2.6
218 GHz Intel Xeon processors, reflecting a typical CTM experimental set-up, using otherwise identical

219 settings and initial chemical conditions as the coupled GEOS-5/GEOS-Chem simulations. Since GEOS-5
220 is a pure MPI application, each MPI process corresponds to a single processor core.

221 Figure 2a gives execution wall times for the total simulation and for the chemistry (GEOS-Chem) and
222 dynamics gridded components. To analyze the performance and scalability results, we define the
223 normalized scaling efficiency S for simulation i as

$$224 \quad 225 \quad S = \left(\frac{W_{x,i-1} - W_{x,i}}{W_{x,i-1}} \right) \left(\frac{N_i}{N_i - N_{i-1}} \right) \quad (4)$$

226 where $W_{x,i}$ is the walltime for component x , and N_i is the number of cores allocated to the simulation. S
227 measures how efficiently the addition of computational resources speeds up execution. For example, a
228 value of 0.9 indicates that a doubling of computational resources decreases walltime by a factor of 1.8. A
229 value of zero means no speed-up. A negative value means slow-down, as might result from increasing
230 I/O. Results shown in Figure 2 for 48 cores are relative to the 8-process GEOS-Chem CTM simulation (i
231 = 0), which uses different shared-memory processes and a different transport code for chemical tracers
232 only. The two simulations are not strictly comparable but results serve to benchmark the performance of
233 the GEOS-5/GEOS-Chem system against the GEOS-Chem CTM.

234 We find that the scaling efficiency for the chemistry module (GEOS-Chem) in the GEOS-5/GEOS-
235 Chem system is close to unity (0.78 ± 0.10) for all numbers of cores, reflecting the independent nature of
236 the chemistry calculation for individual columns. Scaling efficiency of the dynamics and “other”
237 components decreases with increasing number of cores and becomes negative above 192, reflecting the
238 small number of gridpoints allocated to individual cores and hence the increased relative cost of
239 communicating between processes vs. operating within local memory, reflecting what is commonly
240 referred to as weak scaling, or scalability as a function of problem size. Using a large number of cores is
241 less effective for a more coarse resolution simulation.

242 The $0.5^\circ \times 0.625^\circ$ resolution simulation was used to examine the performance of the GEOS-5/GEOS-
243 Chem system when operating on a finer grid resolution than permitted by the GEOS-Chem CTM using
244 shared-memory OpenMP parallelization. The higher resolution also increases the problem size, permitting
245 the efficient use of more computing power. For this simulation, the horizontal grid was decomposed into
246 24×25 lat/lon blocks over 600 cores. The $0.5^\circ \times 0.625^\circ$ resolution simulation completed 0.35 simulation
247 years per wallclock-day.

248 About 20% of the walltime spent on chemistry in the GEOS-5/GEOS-Chem system was spent copying
249 and flipping the vertical dimension of chemical tracer arrays between the GEOS-5 ESM and GEOS-
250 Chem. This would be overcome to a large extent by linking GEOS-Chem tracer arrays to the ESMF using
251 pointers, which access memory locations of preexisting variables directly. This cannot be done within the
252 GEOS-5 ESM for two reasons: (1) GEOS-Chem stores concentrations in double-precision arrays, while
253 the GEOS-5 system generally uses single precision. (2) GEOS-Chem indexes concentration arrays
254 vertically from the surface of the Earth upward while the GEOS-5 system does the reverse. Such
255 limitations are not intrinsic to GEOS-Chem and depend on the specific ESM to which GEOS-Chem is
256 coupled; other ESMs may use different data precision and indexing. Further software engineering in
257 GEOS-Chem could add flexibility in array definitions to accommodate different ESM configurations.

258 Figure 3 illustrates model results with 500 hPa O_3 mixing ratios at 12 UT on July 15, 2006 for GEOS-
259 5/GEOS-Chem simulations at $2 \times 2.5^\circ$ and $0.5^\circ \times 0.625^\circ$ resolutions, and for the GEOS-Chem CTM using
260 GEOS-5 assimilated meteorological data at $2^\circ \times 2.5^\circ$ resolution. All three simulations are initialized from
261

262 the same GEOS-Chem CTM fields at 0 UT on July 1, 2006, but have different meteorology because of
263 differences in resolution and also because the CTM uses assimilated meteorological data while the
264 GEOS-5/GEOS-Chem system in this implementation does not. The Figure demonstrates the fine structure
265 of chemical transport that can be resolved with the $0.5^\circ \times 0.625^\circ$ resolution. The general patterns are
266 roughly consistent between simulations and are reasonable compared to satellite and sonde observations
267 (Zhang et al., 2010). A scatterplot comparing output from the different simulations (Figure 4) shows that
268 they have comparable results. Figures 3 and 4 are intended to illustrate the GEOS-5/GEOS-Chem
269 capability. Quantitative comparison of the GEOS-5/GEOS-Chem and CTM systems will require using the
270 same meteorological data in both, diagnosing the full ensemble of simulated chemical species, and
271 investigating the effect of transport errors when using off-line meteorological fields in the CTM. This
272 comparison is important for investigating the equivalence of the GEOS-Chem ESM module and stand-
273 alone CTM. It represents a major effort and will be documented in a separate publication.

274

275 4. *Summary*

276 We have presented a new grid-independent version of the GEOS-Chem chemical transport model
277 (CTM) to serve as atmospheric chemistry module within Earth system models (ESMs) through the Earth
278 System Modeling Interface (ESMF). The new GEOS-Chem version uses any grid resolution or geometry
279 specified at runtime. The exact same standard GEOS-Chem code (freely available from [http://geos-
280 chem.org](http://geos-chem.org)) supports both ESM and stand-alone CTM applications. This ensures that the continual stream
281 of innovation from the worldwide community contributing to the stand-alone CTM is easily incorporated
282 into the ESM version. The GEOS-Chem ESM module thus always remains state-of-science.

283 We implemented GEOS-Chem as an atmospheric chemistry module within the NASA GEOS-5 ESM
284 and performed a tropospheric oxidant-aerosol simulation (120 coupled chemical species, 66 transported
285 tracers) in that fully coupled environment. Analysis of scalability and performance for 48 to 240 cores
286 shows that the GEOS-Chem atmospheric chemistry module scales efficiently with no degradation as the
287 number of cores increases, reflecting the independent nature of the chemical computation for individual
288 grid columns. Although the inclusion of detailed atmospheric chemistry in an ESM is a major
289 computational expense, it becomes relatively more efficient as the number of cores increases due to its
290 consistent scalability.

291

292

293 **Acknowledgments.** This work was supported by the NASA Modeling, Analysis and Prediction (MAP)
294 Program. The authors thank Ben Auer (NASA-GMAO) and Jack Yatteau (Harvard University) for
295 technical assistance.

References:

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(D19), 23073–23, 2001.

Collins, N., Theurich, G., DeLuca, C., Suarez, M., Trayanov, A., Balaji, V., Li, P., Yang, W., Hill, C. and Silva, A. da: Design and Implementation of Components in the Earth System Modeling Framework, *Int. J. High Perform. Comput. Appl.*, 19(3), 341–350, doi:10.1177/1094342005056120, 2005.

Cotronis, Y. and Dongarra, J.: Recent Advances in Parallel Virtual Machine and Message Passing Interface: 8th European PVM/MPI Users' Group Meeting, Santorini/Thera, Greece, September 23-26, 2001. Proceedings, Springer., 2001.

Eastham, S. D., Weisenstein, D. K. and Barrett, S. R. H.: Development and evaluation of the unified tropospheric–stratospheric chemistry extension (UCX) for the global chemistry-transport model GEOS-Chem, *Atmos. Environ.*, 89, 52–63, doi:10.1016/j.atmosenv.2014.02.001, 2014.

Eller, P., Singh, K., Sandu, A., Bowman, K., Henze, D. K. and Lee, M.: Implementation and evaluation of an array of chemical solvers in the Global Chemical Transport Model GEOS-Chem, *Geosci. Model Dev.*, 2(2), 89–96, 2009.

Fairlie, D. T., Jacob, D. J. and Park, R. J.: The impact of transpacific transport of mineral dust in the United States, *Atmos. Environ.*, 41(6), 1251–1266, doi:10.1016/j.atmosenv.2006.09.048, 2007.

Hairer, E. and Wanner, G.: Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems, Springer., 1996.

Hill, C., DeLuca, C., Suarez, M., da Silva, A. and others: The architecture of the earth system modeling framework, *Comput. Sci. Eng.*, 6(1), 18–28, 2004.

Holmes, C. D., Jacob, D. J., Corbitt, E. S., Mao, J., Yang, X., Talbot, R. and Slemr, F.: Global atmospheric model for mercury including oxidation by bromine atoms, *Atmos Chem Phys*, 10(24), 12037–12057, doi:10.5194/acp-10-12037-2010, 2010.

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, *Atmospheric Chem. Phys.*, 11(7), 3137–3157, doi:10.5194/acp-11-3137-2011, 2011.

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. Discuss.*, 7(1), 1115–1136, doi:10.5194/gmdd-7-1115-2014, 2014.

Koster, R. D., Suarez, M. J., Ducharme, A., Stieglitz, M. and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res. Atmospheres*, 105(D20), 24809–24822, doi:10.1029/2000JD900327, 2000.

Lin, S.-J.: A “vertically Lagrangian” finite-volume dynamical core for global models, *Mon. Weather Rev.*, 132(10), 2293–2307, 2004.

Long, M. S., Keene, W. C., Easter, R., Sander, R., Kerkweg, A., Erickson, D., Liu, X. and Ghan, S.: Implementation of the chemistry module MECCA (v2.5) in the modal aerosol version of the Community

- Atmosphere Model component (v3.6.33) of the Community Earth System Model, *Geosci. Model Dev.*, 6(1), 255–262, doi:10.5194/gmd-6-255-2013, 2013.
- Mao, J., Paulot, F., Jacob, D. J., Cohen, R. C., Crounse, J. D., Wennberg, P. O., Keller, C. A., Hudman, R. C., Barkley, M. P. and Horowitz, L. W.: Ozone and organic nitrates over the eastern United States: Sensitivity to isoprene chemistry, *J. Geophys. Res. Atmospheres*, 118(19), 2013JD020231, doi:10.1002/jgrd.50817, 2013.
- Mirin, A. A. and Worley, P. H.: Improving the performance scalability of the community atmosphere model, *Int. J. High Perform. Comput. Appl.*, 26(1), 17–30, doi:10.1177/1094342011412630, 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. Atmospheres*, 117(D20), D20307, doi:10.1029/2012JD017934, 2012.
- Nassar, R., Jones, D. B. A., Suntharalingam, P., Chen, J. M., Andres, R. J., Wecht, K. J., Yantosca, R. M., Kulawik, S. S., Bowman, K. W., Worden, J. R., Machida, T. and Matsueda, H.: Modeling global atmospheric CO₂ with improved emission inventories and CO₂ production from the oxidation of other carbon species, *Geosci Model Dev Discuss*, 3(3), 889–948, doi:10.5194/gmdd-3-889-2010, 2010.
- Ott, L. E., Bacmeister, J., Pawson, S., Pickering, K., Stenchikov, G., Suarez, M., Huntrieser, H., Loewenstein, M., Lopez, J. and Xueref-Remy, I.: Analysis of Convective Transport and Parameter Sensitivity in a Single Column Version of the Goddard Earth Observation System, Version 5, General Circulation Model, *J. Atmospheric Sci.*, 66(3), 627–646, doi:10.1175/2008JAS2694.1, 2009.
- Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M. and Chin, M.: Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy, *J. Geophys. Res. Atmospheres*, 109(D15), D15204, doi:10.1029/2003JD004473, 2004.
- 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, 5.1.0, and 5.2.0. [online] Available from: <http://ntrs.nasa.gov/search.jsp?R=20120011955> (Accessed 16 June 2014), 2008.
- Sandu, A. and Sander, R.: Technical Note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, *Atmospheric Chem. Phys. Discuss.*, 5(5), 8689–8714, 2005.
- Selin, N. E., Jacob, D. J., Yantosca, R. M., Strode, S., Jaeglé, L. and Sunderland, E. M.: Global 3-D land-ocean-atmosphere model for mercury: Present-day versus preindustrial cycles and anthropogenic enrichment factors for deposition, *Glob. Biogeochem. Cycles*, 22(2), GB2011, doi:10.1029/2007GB003040, 2008.
- Stieglitz, M., Ducharne, A., Koster, R. and Suarez, M.: The Impact of Detailed Snow Physics on the Simulation of Snow Cover and Subsurface Thermodynamics at Continental Scales, *J. Hydrometeorol.*, 2(3), 228–242, 2001.
- Suarez, M., Trayanov, A., da Silva, A. and Chakraborty, P.: MAPL Manual, [online] Available from: http://geos5.org/wiki/images/f/fa/MAPL_UsersGuide.pdf, 2013.

Wang, J. S., Logan, J. A., McElroy, M. B., Duncan, B. N., Megretskaya, I. A. and Yantosca, R. M.: A 3-D model analysis of the slowdown and interannual variability in the methane growth rate from 1988 to 1997, *Glob. Biogeochem. Cycles*, 18(3), GB3011, doi:10.1029/2003GB002180, 2004.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z. and Blake, D. R.: Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, *J. Geophys. Res. Atmospheres*, 119(12), 2014JD021551, doi:10.1002/2014JD021551, 2014.

Wu, S., Mickley, L. J., Jacob, D. J., Logan, J. A., Yantosca, R. M. and Rind, D.: Why are there large differences between models in global budgets of tropospheric ozone?, *J. Geophys. Res. Atmospheres*, 112(D5), D05302, doi:10.1029/2006JD007801, 2007.

Figure 1. Coupling between the GEOS-Chem CTM (dashed beige box) and an ESM (blue box). The schematic shows how the coupling is managed through the ESMF, and utilizes only the GEOS-Chem components bound by the ESM box: Transport modules in the GEOS-Chem CTM are bypassed and replaced by the ESM transport modules through the atmospheric dynamics simulation .

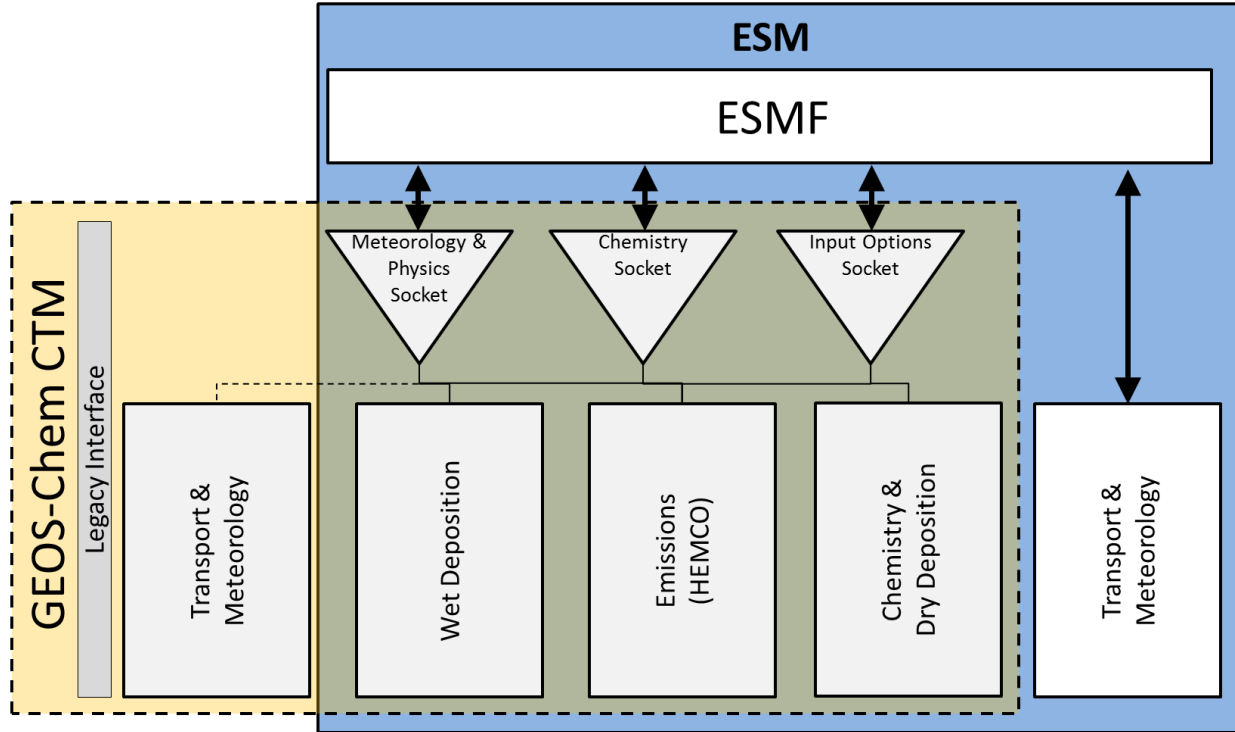


Figure 2. Performance and scalability of the GEOS-5/GEOS-Chem system for a 1-month test simulation including detailed oxidant-aerosol tropospheric chemistry at $2^\circ \times 2.5^\circ$ horizontal resolution. Top panel: total and stacked wall-times for the chemical operator (GEOS-Chem), dynamics, and other routines versus number of processor cores. Bottom panel: Scaling efficiency (Eq. 4) for chemistry, dynamics, and the full GEOS-5/GEOS-Chem system. Values shown for 48 cores are relative to the 8-process shared-memory GEOS-Chem CTM.

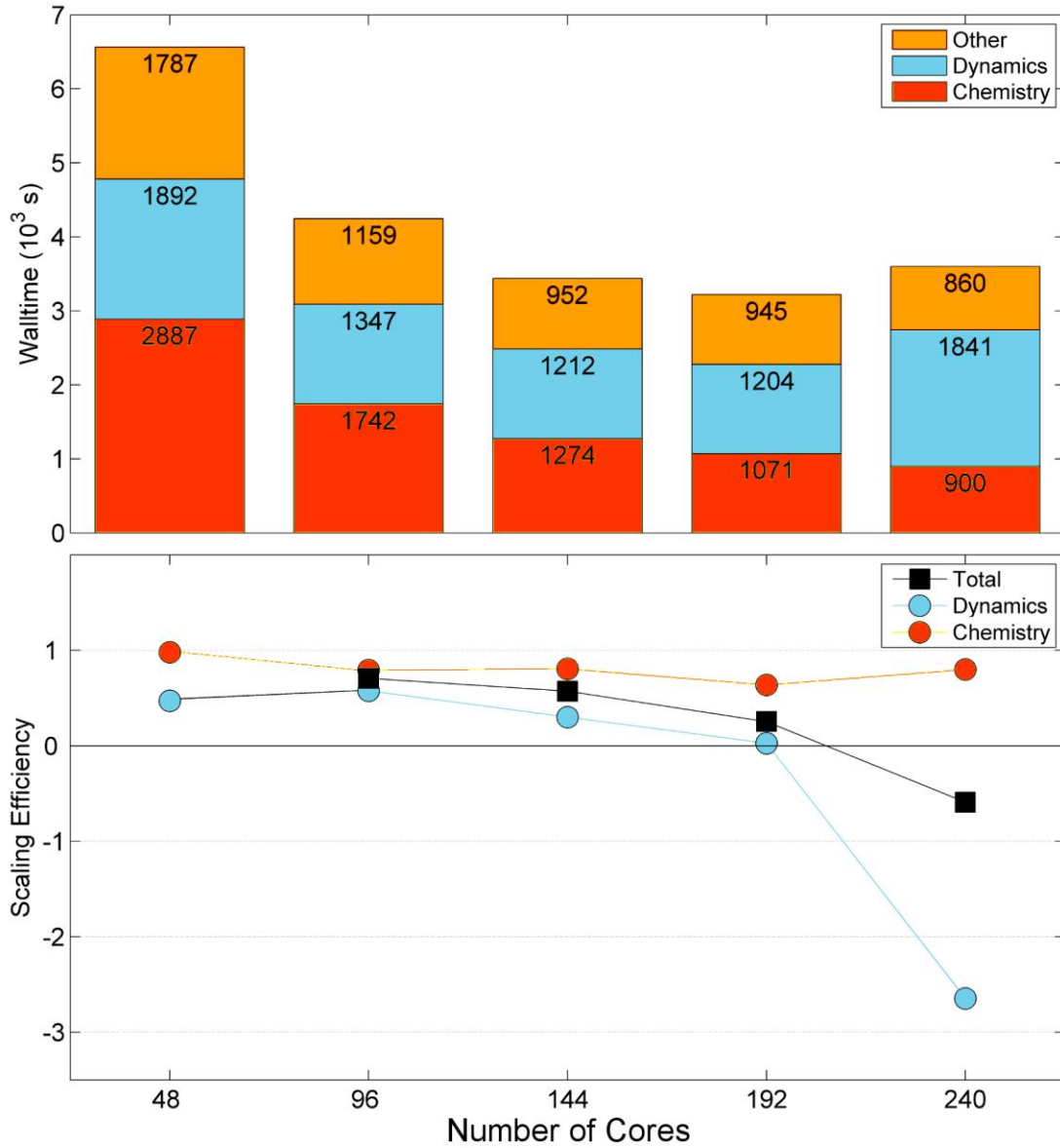


Figure 3. Instantaneous 500 hPa ozone mixing ratios (nmol mol^{-1}) at 12 UT on July 15, 2006, for CTM and ESM implementations of GEOS-Chem. Top panel: GEOS-Chem CTM at $2^\circ \times 2.5^\circ$ resolution driven by GEOS-5 assimilated meteorological data with $0.5^\circ \times 0.67^\circ$ resolution. Middle panel: GEOS-5/GEOS-Chem ESM at $2^\circ \times 2.5^\circ$ resolution. Bottom panel: GEOS-5/GEOS-Chem ESM at $0.5^\circ \times 0.625^\circ$ resolution. All three simulations are initialized with the same GEOS-Chem CTM fields at 0 UT on July 1, 2006, but the ESM as implemented here does not include meteorological data assimilation.

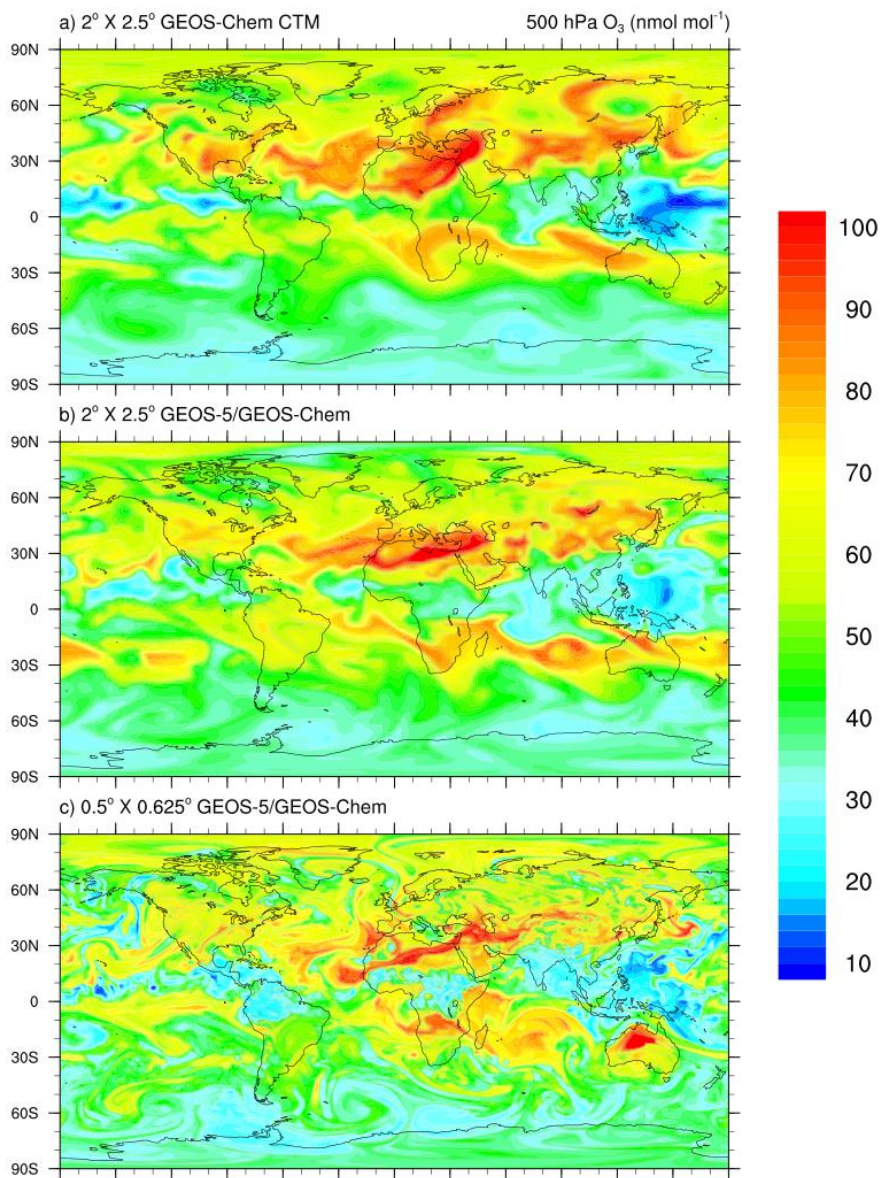


Figure 4. Comparison of instantaneous 500 hPa ozone mixing ratios (nmol mol^{-1}) at 12 UT on July 15, 2006 in the stand-alone GEOS-Chem simulation at $2^\circ \times 2.5^\circ$ horizontal resolution and the coupled GEOS-5/GEOS-Chem simulation at $2^\circ \times 2.5^\circ$ (red) and $0.5^\circ \times 0.625^\circ$ (blue) resolutions. The $0.5^\circ \times 0.625^\circ$ results are regridded to $2^\circ \times 2.5^\circ$ resolution, and each point represents a $2^\circ \times 2.5^\circ$ grid square. The reduced-major-axis regression parameters and the 1:1 line are also shown.

