



Development and Application of Coupled Atmospheric Chemistry Models

Citation

Lin, Haipeng. 2024. Development and Application of Coupled Atmospheric Chemistry Models. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37378720>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), 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](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Harvard John A. Paulson School of Engineering and Applied Sciences
have examined a dissertation entitled:

“Development and Application of Coupled Atmospheric Chemistry Models”

presented by: Haipeng Lin

Signature *Daniel Jacob*
Typed name: Professor Daniel J. Jacob

Signature *Scot Martin*
Typed name: Professor Scot Martin

Signature *Kaighin McColl*
Typed name: Professor Kaighin McColl

Signature *Steven C Wofsy*
Typed name: Professor Steve Wofsy

April 12, 2024

Development and Application of Coupled Atmospheric Chemistry Models

A DISSERTATION PRESENTED
BY
HAIPENG LIN
TO
THE SCHOOL OF ENGINEERING AND APPLIED SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
ENVIRONMENTAL SCIENCE & ENGINEERING
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
APRIL 2024

©2024 HAIPENG LIN
ALL RIGHTS RESERVED.

Development and Application of Coupled Atmospheric Chemistry Models

ABSTRACT

Comprehensive and accurate representation of atmospheric chemistry in Earth System Models is a major priority in geoscientific modeling. GEOS-Chem is a state-of-the-science atmospheric chemistry model used by hundreds of research groups worldwide. It has been restructured to operate as a chemical module in weather and climate models while sharing the same scientific codebase as the standalone chemical transport model. The Community Earth System Model version 2 (CESM2) is a state-of-science Earth System Model that has participated in multiple model intercomparison activities. CESM2 includes a comprehensive chemistry option, CAM-chem, with a different development heritage from GEOS-Chem. The two models show large differences in model skill in comparison to observations in different regions, implying differences in sensitivity to perturbations. Previous model intercomparisons generally compared entire modeling systems, introducing first-order differences complicating the attribution of particular strengths in reproducing observations to specific representations of processes in each model. Implementation of GEOS-Chem within CESM2 as an alternative chemistry option to CAM-chem allows for detailed process-based intercomparison between two models, informing more accurate representation of atmospheric chemistry, and contributing to the vision of a modular chemistry-climate model.

In this work, we develop a multi-model emissions and data tool through an updated version of the Harmonized Emissions Component (HEMCO) and couple it to the CESM2 model. This serves as the foundation for implementation of GEOS-Chem within CESM2 and intercompari-

son with CAM-chem (Chapter 1). Because of the increasing chemical complexity in atmospheric chemistry models resulting in increased computational burden, which is particularly relevant in coupled model environments, we develop an adaptive, auto-reduction solver for chemical kinetics that can reduce the complexity of the chemical mechanism depending on local conditions. It can provide a $\sim 30\%$ speed-up in the chemical solver while introducing $\sim 1\%$ error in the troposphere. We evaluate the adaptive solver in the offline GEOS-Chem chemical transport model at a $2^\circ \times 2.5^\circ$ resolution, but it has been implemented in the Kinetic Pre-Processor (KPP) software package making it easily extensible beyond GEOS-Chem (Chapter 2). We then apply the coupled chemistry-climate model GEOS-Chem within CESM2 to compare its representation of tropospheric oxidant chemistry against the default CAM-chem chemistry driven by the same dynamics, physics, and emissions. We use ozonesondes and aircraft observations from the ATom-1 and KORUS-AQ campaigns to identify and attribute major differences between the two models. (Chapter 3).

Contents

TITLE PAGE	i
COPYRIGHT	ii
ABSTRACT	iii
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
PRIOR PUBLICATIONS AND CONTRIBUTING AUTHORS	x
ACKNOWLEDGMENTS	xi
1 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	I
1.1 Introduction	3
1.2 HEMCO 3.0 description	6
1.3 Implementations of HEMCO 3.0 in different models	19
1.4 Conclusions	30
1.5 Code and data availability	32
1.6 Author contributions and acknowledgments	32
2 AN ADAPTIVE AUTO-REDUCTION SOLVER FOR SPEEDING UP INTEGRATION OF CHEMICAL KINETICS IN ATMOSPHERIC CHEMISTRY MODELS: IMPLEMENTATION AND EVALUATION IN THE KINETIC PRE-PROCESSOR (KPP) VERSION 3.0.0	34
2.1 Introduction	35
2.2 Adaptive Solver for Chemical Kinetics	37
2.3 Kinetic Pre-Processor (KPP) Version 3.0.0	41

2.4	Adaptive Mechanism Auto-Reduction in the GEOS-Chem Model Using KPP Ver- sion 3.0.0	49
2.5	Conclusions	56
2.6	Code and data availability	57
2.7	Acknowledgments	58
3	INTERCOMPARISON OF GEOS-CHEM AND CAM-CHEM TROPOSPHERIC OXIDANT CHEMISTRY WITHIN THE COMMUNITY EARTH SYSTEM MODEL VERSION 2	59
3.1	Introduction	60
3.2	Model description and methods	63
3.3	Comparison of photolysis schemes: Fast-JX and TUV	67
3.4	Global budgets and distributions of tropospheric oxidants	70
3.5	Comparisons to ozonesondes and to ATom-1 aircraft campaign	75
3.6	Comparison to KORUS-AQ aircraft campaign	79
3.7	Conclusions	81
3.8	Code availability	84
3.9	Author contributions and Acknowledgements	84
	APPENDIX A APPENDIX A: SUPPORTING INFORMATION FOR CHAPTER 2	85
	REFERENCES	92

Listing of figures

1.1	Three-layer architecture of HEMCO 3.0 including the Data Input Layer, the HEMCO Core, and the Model Interface Layer.	7
1.2	Sample HEMCO configuration file.	9
1.3	Benefit of using a HEMCO grid finer than the model grid for masking of emissions by a regional inventory.	17
1.4	HEMCO 3.0 implementation in the GEOS-Chem Classic model.	20
1.5	HEMCO 3.0 implementation in the GEOS-Chem High-Performance model (GCHP).	21
1.6	HEMCO 3.0 implementation in the NASA GEOS Earth System Model.	22
1.7	HEMCO 3.0 implementation in the WRF-GC model.	24
1.8	HEMCO 3.0 implementation in CESM2.	26
1.9	Planned HEMCO 3.0 implementation in the NOAA UFS.	28
2.1	Example KPP input file.	42
2.2	Mechanism auto-reduction in KPP.	47
2.3	Percentage of GEOS-Chem species retained in the fast sub-mechanism when using the adaptive solver.	51
2.4	Accuracy of the adaptive solver in a 1-year GEOS-Chem simulation.	53
2.5	Relative error of the adaptive solver for ozone, OH, and NO ₂ globally.	54
3.1	Mean photolysis frequencies from Fast-JX and differences to TUV for NO ₂ (J_{NO_2}) and O ₃ to O(¹ D) (J_{O1D}) in surface air in July 2016	67
3.2	Median vertical profiles of J_{NO_2} and J_{O1D} from the KORUS-AQ aircraft campaign over the Seoul Metropolitan Area.	69
3.3	Annual mean OH concentrations in surface air and at 500 hPa in GEOS-Chem within CESM2, and differences with CAM-chem.. . . .	71
3.4	Annual mean ozone and NO _x concentrations in surface air and at 500 hPa in GEOS-Chem within CESM2, and differences with CAM-chem.	73
3.5	Comparison of GEOS-Chem and CAM-chem simulated annual mean ozone vertical profiles to 2016 ozonesonde observations.	76

3.6	Median vertical profiles of OH, NO, and CO concentrations from the ATom-1 field campaign (July-August 2016) and from the GEOS-Chem and CAM-chem models within CESM2.	77
3.7	Median tropospheric vertical profiles of species concentrations over the Seoul Metropolitan Area during the KORUS-AQ aircraft campaign in May-June 2016.	80
A.1	Mass imbalance for total odd oxygen, reactive nitrogen, chlorine, and bromine atoms.	87
A.2	Accuracy of the adaptive solver in a 1-year GEOS-Chem simulation, keeping halogen species active.	90
A.3	Absolute error of the adaptive solver for ozone, OH, and NO_2 globally.	91

TO THOSE WHO HAVE SELFLESSLY DEDICATED THEMSELVES
TO FURTHERING THE ACADEMIC ENTERPRISE.

”LUMINOUS BEINGS ARE WE, NOT THIS CRUDE MATTER.” – MASTER YODA

Prior Publications and Contributing Authors

The following authors contributed to Chapter 1: Daniel J. Jacob, Elizabeth W. Lundgren, Melissa P. Sulprizio, Christoph A. Keller, Thibaud M. Fritz, Sebastian D. Eastham, Louisa K. Emmons, Patrick C. Campbell, Barry Baker, Rick D. Saylor, and Raffaele Montuoro.

It has previously appeared as:

Lin, H., Jacob, D. J., Lundgren, E. W., Sulprizio, M. P., Keller, C. A., Fritz, T. M., Eastham, S. D., Emmons, L. K., Campbell, P. C., Baker, B., Saylor, R. D., and Montuoro, R.: *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*, Geosci. Model Dev., 14, 5487–5506, doi:10.5194/gmd-14-5487-2021, 2021.

The following authors contributed to Chapter 2: Michael S. Long, Rolf Sander, Adrian Sandu, Robert M. Yantosca, Lucas A. Estrada, Lu Shen, and Daniel J. Jacob.

It has previously appeared as:

Lin, H., Long, M. S., Sander, R., Sandu, A., Yantosca, R. M., Estrada, L. A., Shen, L., and Jacob, D. J.: *An adaptive auto-reduction solver for speeding up integration of chemical kinetics in atmospheric chemistry models: Implementation and evaluation in the Kinetic Pre-Processor (KPP) version 3.0.0*, J. Adv. Model. Earth Sy., 15, e2022MS003293, doi:10.1029/2022MS003293, 2023.

The following authors contributed to Chapter 3: Louisa K. Emmons, Elizabeth W. Lundgren, Laura Hyesung Yang, Xu Feng, Ruijun Dang, Shixian Zhai, Yunxiao Tang, Makoto M. Kelp, Nadia K. Colombi, Sebastian D. Eastham, Thibaud M. Fritz, and Daniel J. Jacob.

It is currently under consideration for publication in *Atmos. Chem. Phys.*

Acknowledgments

I am grateful to my advisor, Daniel Jacob. As "the father of GEOS-Chem", he has shaped the very foundation that my research is built upon, and taught a software engineer how to be a scientist.

To Loretta Mickley, who has provided extensive advice and feedback with great kindness.

To my dissertation committee, Steven Wofsy, Scot Martin, and Kaighin McColl, for their guidance and support throughout this journey.

I am grateful to my mentor at NCAR, Louisa Emmons, for her kindness in hosting me throughout many projects and for passing on invaluable knowledge on CESM, CAM-chem, and atmospheric chemistry in general.

Thank you to my mentors in the ACMG, Zichong Chen, Zhen Qu, Hannah Nesser, and Jiawei Zhuang, who have helped me navigate the ups and downs of a graduate student. Thank you to Christoph Keller, Michael Long, Thibaud Fritz, and Sebastian Eastham, whose work formed the foundation for my research projects.

My work would not have been possible without my collaborators and peers - Viral Shah, Kelvin Bates, Ruijun Dang, Xu Feng, Yujin Oak, Makoto Kelp, Nadia Colombi, Ellie Beaudry, and Laura Yang from the air quality and chemistry subgroup; Robert Yantosca, Melissa Sulprizio, Lizzie Lundgren, and Lucas Estrada in the GEOS-Chem Support Team; and Duseong Jo and Wenfu Tang at NCAR. Their valuable feedback and productive discussions shaped my research.

I am indebted to the alumni and members of the ACMG who shape this group and carry its culture. In this group I have met bright minds and kind hearts with whom I hope to forge lifelong friendships with. Special thanks to Zichong, Ruijun, Xu, James, Yujin, Xiaolin, Elise, Makoto, Nadia, Jack, Ellie, Laura, Sarah, Nick, Danyuting, and Yunxiao, with whom I shared many hours of blood, sweat, tears, but also joy.

To my family who have always supported me and loved me unconditionally.

To my grandmother, a teacher who has loved and inspired all those around her and has given everything she had to her family for her whole life.

To my wife, Jingyun, for her unwavering support, understanding, and love. Jingyun has kept the lights on in our small family, literally and figuratively, ever since the beginning of my research career with long nights developing WRF-GC in the spring of 2018.

During the pandemic, she fearlessly traveled thousands of miles overcoming closed borders and consulates - twice - so that we could be together. Her countless efforts of resolution and courage forever motivates me, and as I start a new journey after the Ph.D., it is time I do the same.

1

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

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 algorithms. 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 offline stand-alone mode, but more importantly it provides an online facility for models to compute emissions at runtime. HEMCO complies with the Earth System Modeling Framework (ESMF) for portability across models. We present a new version here, 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) the 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-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 or vegetation). 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), as well as the planned implementation in the NOAA Unified Forecast System (UFS). Implementation of HEMCO in CESM2 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.

1.1 INTRODUCTION

Emissions are a crucial component in modeling 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, 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 offline or online. An offline emissions module precomputes emissions on the target model grid and archives them as time-varying files for input to the model. An online 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. Offline processing of emissions is used by many models, such as the PREP-CHEM-SRC preprocessor system (Freitas et al., 2011) in 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/>, last access: 2 September 2021) in the CMAQ (Byun and

Schere, 2006) and CAMx (Ramboll Environment and Health, 2020) models. The advantage of computing emissions at the preprocessing stage is the versatility in preprocessing tools and low computational requirements at runtime. However, the preprocessing 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 preprocessor code. Emissions dependent on environmental variables computed in the atmospheric model cannot be treated offline, complicating the infrastructure.

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 offline in a stand-alone mode to compute and archive emissions or online 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. 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.

HEMCO was originally developed for the GEOS-Chem atmospheric chemistry model (Bey et al., 2001; Eastham et al., 2018), wherein 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 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) uses a different implementation of 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 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 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 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 multiscale 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 multiscale 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.

1.2 HEMCO 3.0 DESCRIPTION

1.2.1 GENERAL ARCHITECTURE

HEMCO 3.0 is modularized into a three-layer architecture as shown in Figure 1.1, consisting of the Data Input Layer, the HEMCO Core, and the Model Interface Layer. The Data Input Layer reads the configuration file and the database library of emission inventories and other environmental information, and it 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 including adding, scaling, and masking of individual inventories, as well as computing emissions dependent on model state variables and environmental data (through algorithms referred to as HEMCO extensions). The Model Interface Layer communicates HEMCO output (including emission fluxes, 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

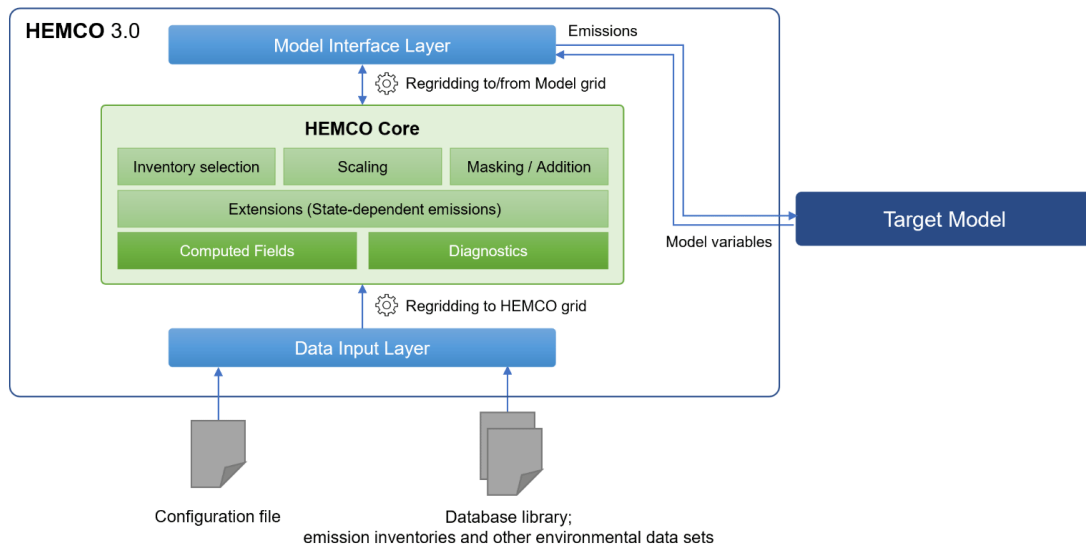


Figure 1.1: Three-layer architecture of HEMCO 3.0 including the Data Input Layer, the HEMCO Core, and the Model Interface Layer.

Model 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 multiscale model grid), with the other layers handling the re-gridding to and 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 databases (such as for aircraft emissions) are re-gridded vertically to the model grid through the Data Input Layer.

The HEMCO configuration file (example in Figure 1.2) 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 datasets that is updated with every new GEOS-Chem version release, but users can readily add their own by 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, last access: 7 January 2021).

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 preprocessing 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 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 Figure 1.2, 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

(1) Collection switch	<pre> ### BEGIN SECTION EXTENSION SWITCHES 0 Base : on * --> CEDS : true --> MEIC : true ### END SECTION EXTENSION SWITCHES ### </pre>
(2) Data containers	Data container name Data source (File name, variable, temporal range and refresh frequency, cycling option, spatial dimensions) Units and model species ("*" if not emissions data) Category and hierarchy <pre> ### BEGIN SECTION BASE EMISSIONS (((CEDS 0 CEDS_EOH_AGR CEDS_EOH_YYYY.nc VOC1 1970-2017/1-12/1/0 C xy kg/m2/s CH3OH 91 1 1 0 CEDS_EOH_AGR CEDS_EOH_YYYY.nc VOC1 1970-2017/1-12/1/0 C xy kg/m2/s C2H5OH 92 1 1 0 CEDS_NO_AGR CEDS_NO_YYYY.nc NO_agr 1970-2017/1-12/1/0 C xy kg/m2/s NO 25 1 1 0 CEDS_NO_IND CEDS_NO_YYYY.nc NO_ind 1970-2017/1-12/1/0 C xyL* kg/m2/s NO 25/316 1 1)))CEDS (((MEIC 0 MEIC_NO_AGR MEIC_NO.05x0666.nc NO_agri 2000-2017/1-12/1/0 C xy kg/m2/s NO 301/1009 1 2)))MEIC </pre> <i>CEDS Collection</i> <i>MEIC Collection</i>
(3) Scaling factors and masks	<pre> * UV_ALBEDO uvalbedo.geos.2x25.nc UVALBD 1985/1-12/1/0 C xy percent * - 1 1 ### END SECTION BASE EMISSIONS ### ### BEGIN SECTION SCALE FACTORS 25 EDGAR_TODNOX EDGAR_hourly_NOxScal.nc NOxscale 2000/1/1/* C xy unitless 1 301 MEIC_DOM_AGR 0.994/1.001/1.001/1.001/1.001/1.001 g_industry 2017/1/1/0 C xyz 1 316 INDUSTRY_LEVS vert_alloc.nc - - - xy 1 1 91 VOC1toCH3OH 0.15 - - - xy 1 1 92 VOC1toC2H5OH 0.85 - - - xy 1 1 ### END SECTION SCALE FACTORS ### ### BEGIN SECTION MASKS 1009 CHINA_MASK China_mask.generic.1x1.nc MASK 2000/1/1/0 C xy 1 1 70/10/130/60 ### END SECTION MASKS ### </pre>

Figure 1.2: 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).

provided as kilograms of NO_x on an NO₂ mass basis in the inventory file but emitted as NO and NO₂ in the model. Scaling factors can be specified for each individual entry in the HEMCO configuration file, allowing 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.

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.

1.2.2 HEMCO CORE

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.

Table 1.1 lists the emission inventories currently in the HEMCO 3.0 default database library. Users can select from that list and 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

Table 1.1: Emission inventories currently in the default HEMCO 3.0 library.

Coverage	Species ^a	Resolution (lat × long)	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 ₂ H ₅ OH	2° × 2.5°	Millet et al. (2010), Olivier et al. (2003), Granier et al. (2005)
Global	Fossil fuel and biofuel C ₂ H ₆	4° × 5°	Tzompa-Sosa et al. (2017)
Global	C ₃ H ₈	2° × 2.5°	Xiao et al. (2008)
Global	CHBr ₃ , CH ₂ Br ₂	2° × 2.5°	Liang et al. (2010)
Global	CH ₃ I, CH ₂ I ₂ , CH ₂ ICl, CH ₂ IBr	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 ₃ CHO, C ₂ H ₅ OH	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 ^b	Fire CO, NO _x , NH ₃ , SO ₂ , VOCs, BC, OC	0.25° × 0.25°	GFED v4.1 (Akagi et al., 2011; Andreae and Merlet, 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 and da Silva, 2015)
Global	Fire CO, NO _x , NH ₃ , SO ₂ , VOCs, BC, OC	0.25° × 0.25°	BB4CMIP (van Marle 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 access: 23 July 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 access: 23 July 2021)
Africa	CO, NO _x , NH ₃ , SO ₂ , VOCs, BC, OC	0.1° × 0.1°	DICE-Africa (Marais and Wiedinmyer, 2016)
China	CO, NO _x , NH ₃ , SO ₂ , VOCs, BC, OC	0.5° × 0.66°	MEIC (http://meicmodel.org/ , last access: 23 July 2021; Li et al., 2017; Zheng et al., 2018)

^a VOCs: volatile organic compounds; OC: organic carbon aerosol; BC: black carbon aerosol; ^b implemented as an extension to distribute dry matter input data into model species.

Table 1.2: Emission extensions available in HEMCO 3.0 as built-in algorithms.

Species*	Extension name and reference
Oceanic DMS, acetone, acetaldehyde, methyl nitrate, ethyl nitrate, methanol	SeaFlux (Johnson, 2010)
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., 2002; 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)

* DMS: dimethyl sulfide.

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 and specify the region that the inventory is used for and its precedence over the existing entries in the HEMCO library without necessitating preprocessing of the inventory source files.

Emissions dependent on model state such as dust or lightning can be computed online 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 precompute these emissions based on offline input data and disable the HEMCO extension. Both approaches are routinely used in GEOS-Chem (Weng et al., 2020). Table 1.2 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.

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. Emission fluxes are summed by species, and non-emissions data are stored individually by their data container name (UV albedo example in Figure 1.2). 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 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 the purpose of 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 47-level GEOS grid when reading vertically distributed emissions. This required preprocessing of the emission inventory files from their original vertical grids to the sup-

ported GEOS grid. Such preprocessing is no longer required 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 online 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 Figure 1.2 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 3.16 in Figure 1.2. 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 can thus be applied to an individual inventory, sector, or species. Detailed documentation of this capability is available in the HEMCO user guide (http://wiki.seas.harvard.edu/geos-chem/index.php/HEMCO_examples#Applying_2D_emissions_vertically, last access: 8 August 2021).

HEMCO also includes as a diagnostic capability a NetCDF output component to archive selected emissions at specified time 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 offline in stand-alone mode and also in GEOS-Chem Classic wherein 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.

1.2.3 DEFAULT DATA INPUT LAYER AND MODEL INTERFACE LAYER

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.

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 that use a non-rectilinear grid for computation or data, it would be necessary to modify the default Data Input Layer if re-gridding is to be 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 interface layer to provide HEMCO with information on the model environment and update 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. 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 config-

uration 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 CESM, as described in Sect. 3.5.

1.2.4 HEMCO GRID

HEMCO 3.0 provides the ability to compute emissions online on a horizontal grid finer than the model grid. Previous versions of HEMCO assumed its operation to be 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 at a higher resolution for masking and scaling purposes, thus achieving greater accuracy at boundaries between different inventory domains, and (3) use high-resolution environmental datasets when computing emissions through extensions.

Figure 1.3 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 Figure 1.3, 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

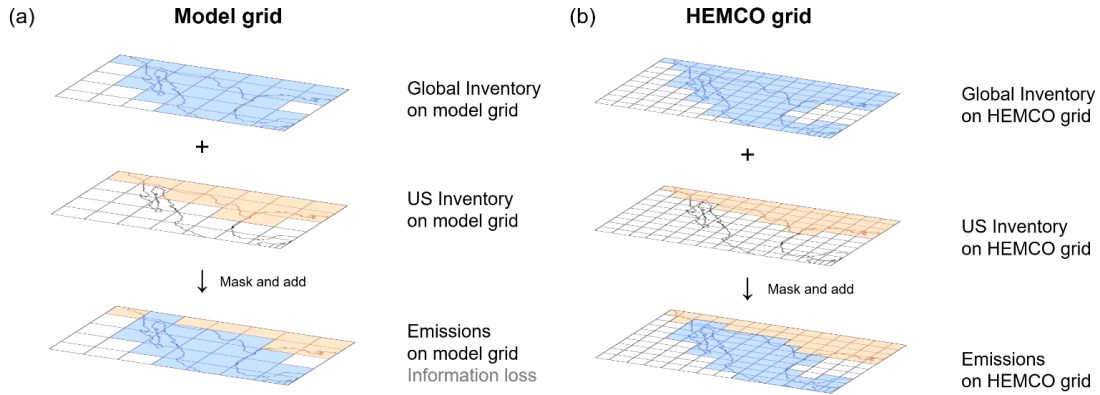


Figure 1.3: Benefit of using a HEMCO grid finer than the model grid for masking of emissions by a regional inventory. The example in the figure illustrates a case in which a US inventory (orange) is used to replace a global default inventory (blue) in a model simulation on a coarse $4^\circ \times 5^\circ$ 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 finer ($2^\circ \times 2.5^\circ$) 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.

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 nonlinear 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 $4^\circ \times 5^\circ$ or $2^\circ \times 2.5^\circ$ we have found that a HEMCO grid of $1^\circ \times 1.25^\circ$ 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 preprocessing the emissions on their native grids using HEMCO in offline mode. Another option is to use a regional rather than global model. For example, in nested GEOS-Chem Classic simulations at $0.25^\circ \times 0.3125^\circ$ resolution we find that HEMCO can easily handle a HEMCO grid of $0.1^\circ \times 0.1^\circ$, which is 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 in which 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 online re-gridding is implemented in the Model Interface Layer.

1.2.5 DATA BROKER FUNCTIONALITY

HEMCO 3.0 has the capability to process any model input data other than emissions such as meteorological fields, land use maps, and boundary conditions. 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 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.

1.3 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, 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.

1.3.1 GEOS-CHEM "CLASSIC"

GEOS-Chem Classic (Bey et al., 2001) is an offline 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 1.4 shows the implementation of HEMCO 3.0 in 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

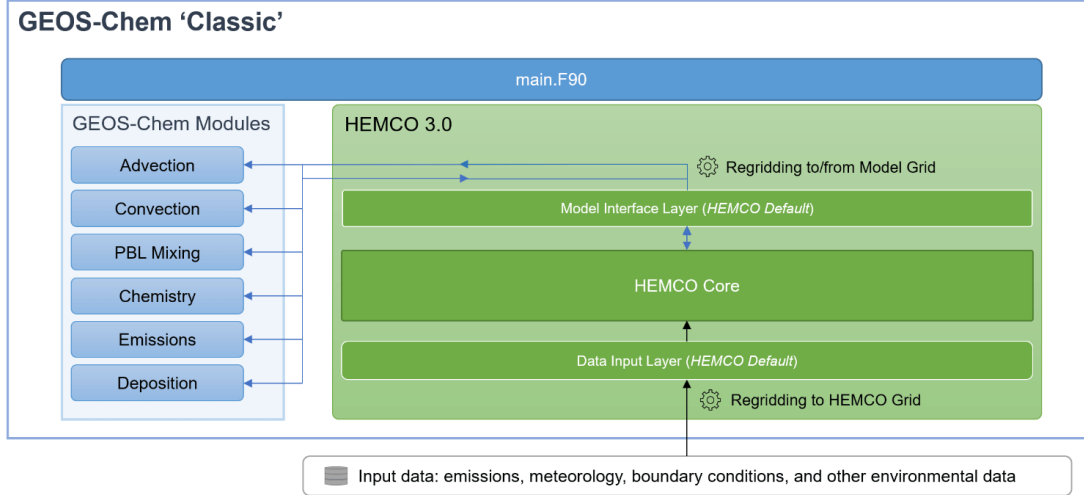


Figure 1.4: HEMCO 3.0 implementation in the GEOS-Chem Classic model. HEMCO in GEOS-Chem Classic is a general input data broker. The main model driver routine `main.F90` successively calls each GEOS-Chem module and HEMCO in a time step loop. HEMCO 3.0 reads all input data through the default Data Input 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 modules that are used by HEMCO to compute state-dependent emissions through extensions.

horizontal re-gridding between the two grids for the data flowing through it.

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 datasets 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 wherein regional nested domains are selected at runtime and all input data are processed for these domains (Li et al., 2021; Shen et al., 2021).

1.3.2 GEOS-CHEM HIGH PERFORMANCE (GCHP)

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

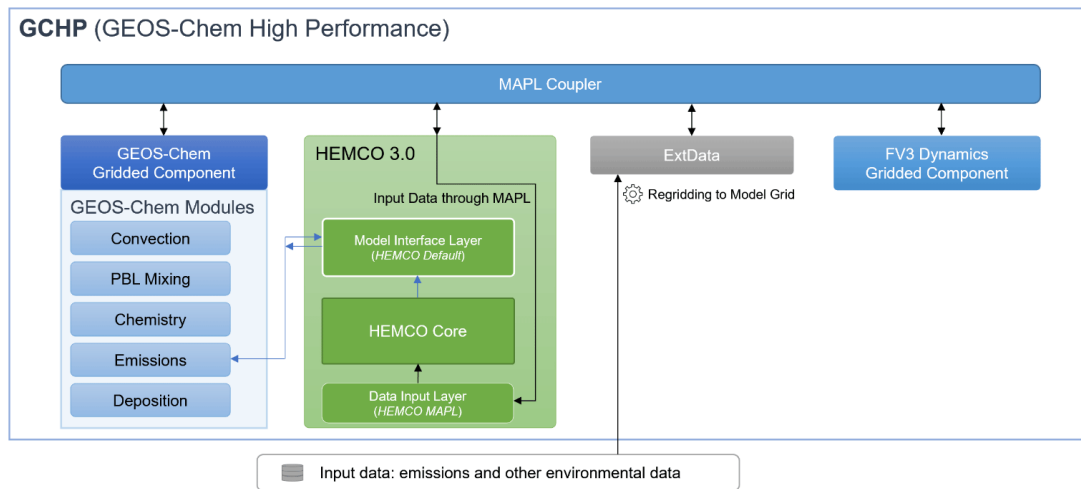


Figure 1.5: HEMCO 3.0 implementation in the GEOS-Chem High-Performance model (GCHP). GCHP operates under the MAPL coupling framework in an MPI parallel environment. The ExtData component reads all data into MAPL, re-gridding them to the model grid, and the input data are retrieved by the HEMCO MAPL implementation 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.

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 1.5 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

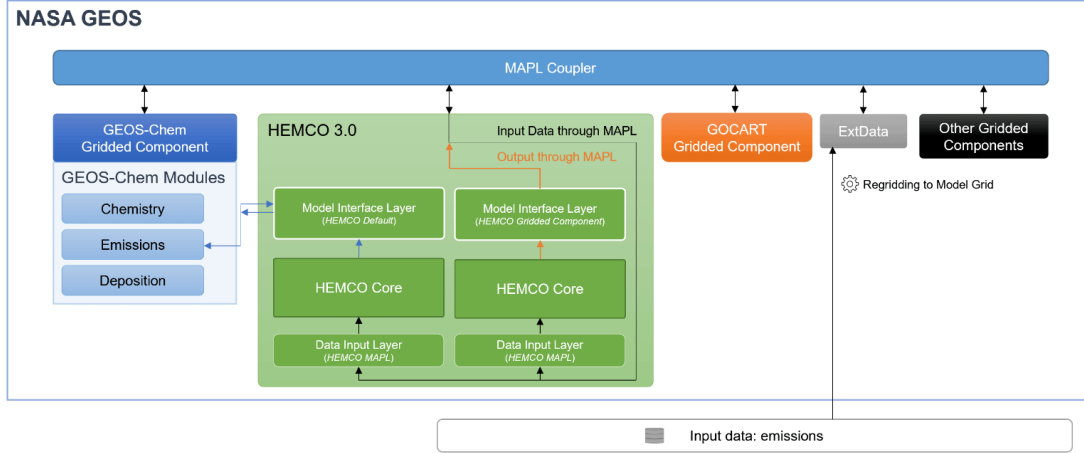


Figure 1.6: HEMCO 3.0 implementation in the NASA GEOS Earth System Model. GEOS is driven by the MAPL framework in an MPI parallel environment and is composed of a number of gridded components for the different ESM operators. GOCART provides a fast aerosol simulation and GEOS-Chem provides detailed chemistry. 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.

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.

1.3.3 NASA GEOS ESM

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

Figure 1.6 shows the implementation of HEMCO 3.0 in the GEOS ESM to serve both the GEOS-Chem and GOCART modules. 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 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.

1.3.4 WRF-GC

The WRF-GC online model (Weather Research and Forecasting Model with GEOS-Chem chemistry; Lin et al., 2020; Feng 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

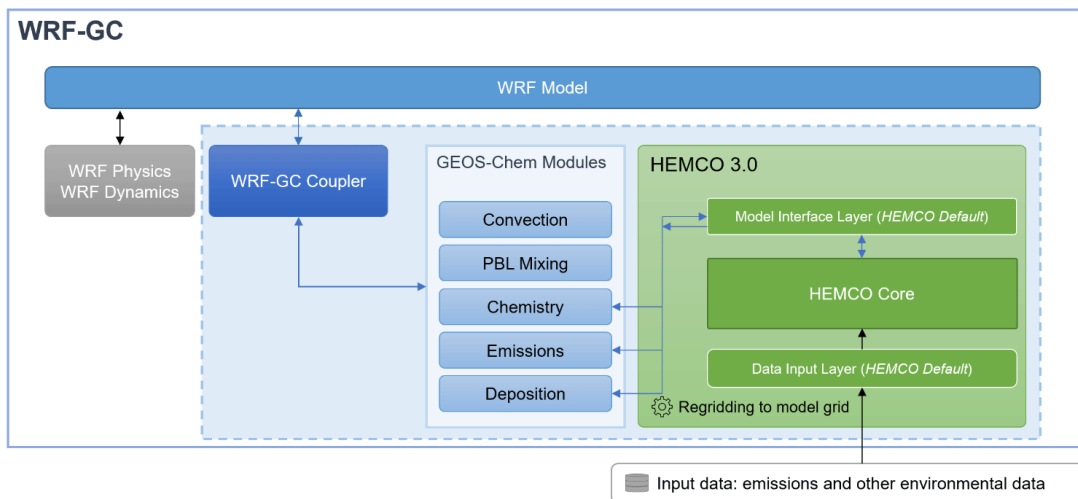


Figure 1.7: HEMCO 3.0 implementation in the WRF-GC model. The model is driven by WRF, which calls GEOS-Chem chemistry through the WRF-GC coupler. HEMCO 3.0 is implemented entirely as part of GEOS-Chem in WRF-GC, with each successive domain in the WRF model containing a separate instance of GEOS-Chem 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.

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 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. Multiscale WRF grids communicating by two-way nesting are also supported by WRF-GC.

Figure 1.7 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.

1.3.5 COMMUNITY EARTH SYSTEM MODEL VERSION 2 (CESM2)

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.

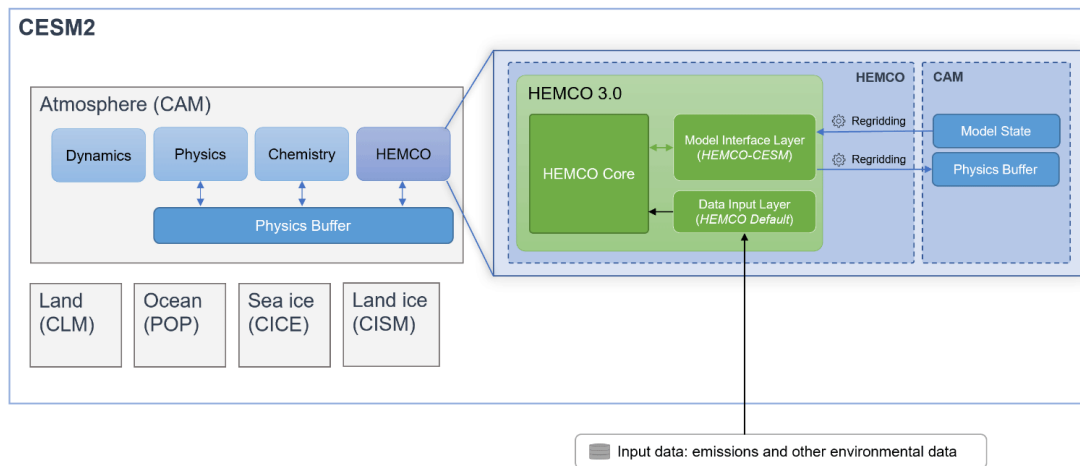


Figure 1.8: HEMCO 3.0 implementation in CESM2. HEMCO 3.0 lives as a component within the CESM2 atmosphere (CAM), reading data through the HEMCO default Data Input Layer and communicating with other CESM components through the HEMCO–CESM implementation of the Model Interface Layer, which reads necessary 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 and 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.

Figure 1.8 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 online re-gridding (<https://earthsystemmodeling.org/regrid/>, last access: 2 September 2021) 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.

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 implemen-

tation 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.

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, wherein atmospheric chemistry affects ecosystem state (both directly and indirectly through the climate), which in turn affects atmospheric chemistry. Alternatively, one may want to use independent environmental data specified 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.

1.3.6 NOAA GEFS-AEROSOL AND NOAA UFS

As described in NOAA’s 2018 Strategic Implementation Plan for next-generation modeling systems (https://www.weather.gov/sti/stimodeling_nggps_implementation, last access: 2 September 2021), “a unified emission system with the capability of providing model-ready, global anthropogenic and natural source emissions inputs for aerosol and gas-phase atmospheric composition across scales is needed.” The next-generation modeling systems of NOAA are being realized as the Unified Forecast System (UFS; <https://ufscommunity.org>, last access: 2 September 2021), 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 Emissions and eXchange Unified System (NEXUS)

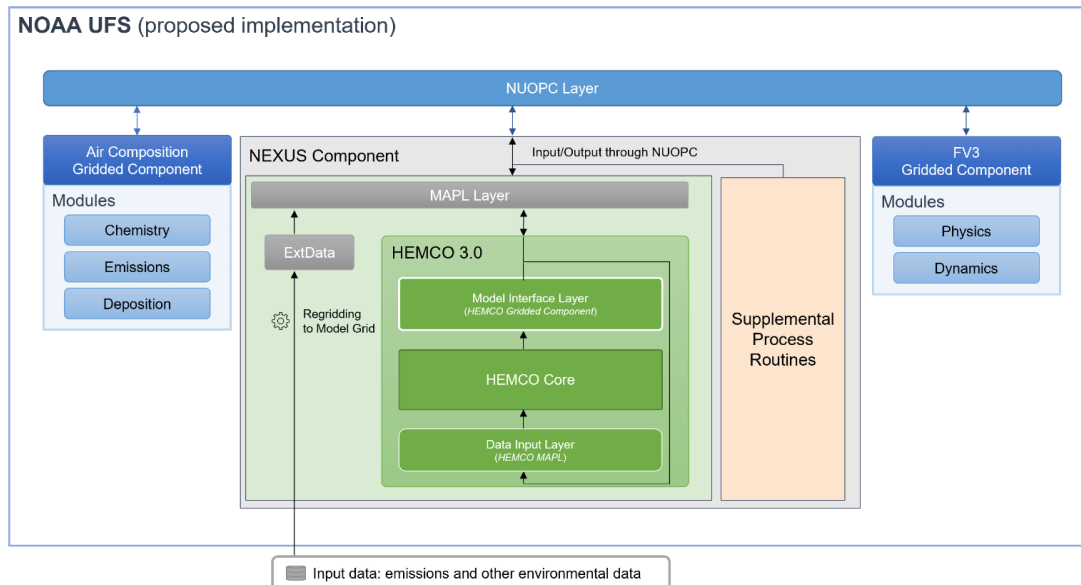


Figure 1.9: Planned HEMCO 3.0 implementation in the NOAA UFS. UFS is a unified software infrastructure for NOAA forecast models to enable exchange of model components between different application models. HEMCO is planned as the core of the NOAA Emissions and eXchange Unified System (NEXUS) component 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 (Figure 1.6) but reads independent environmental data for use in surface exchange calculations. Supplemental process routines provide the surface exchange calculations. NEXUS operates under the NUOPC coupling framework to provide emissions and surface exchange information to other UFS components.

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 1.9 shows the planned implementation of HEMCO as the core of NEXUS within the NOAA UFS. Currently, HEMCO is used as an offline emissions preprocessor 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 that uses the NASA GOCART aerosol model; (ii) an online, regional-scale air quality model with fully coupled gas-aerosol chemistry derived from the Community Multiscale Air Quality (CMAQ) model (United States Environmental Protection Agency, 2018); and (iii) an online, regional-scale rapid refresh forecast system (RRFS) that includes fully coupled smoke and dust emissions as well as transport and aerosol–weather interactions. Further planned development of NEXUS for UFS AAC models includes the integration of HEMCO 3.0 as a dynamic online emissions processor for both anthropogenic inventories and natural, process-based sources (e.g., windblown dust, wildfire smoke, sea salt). 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 es-

established coupling and integration infrastructures, such as the National Unified Operational Prediction Capability (NUOPC) layer (Theurich et al., 2016; <http://earthsystemmodeling.org/nuopc/>, last access: 2 September 2021).

1.4 CONCLUSIONS

We presented an updated version 3.0 of the Harmonized Emissions Component (HEMCO 3.0) for atmospheric models. HEMCO is a versatile online 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 datasets 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), CESM2 with either CAM-chem or GEOS-Chem chemistry, and the NOAA GEFS-Aerosol model as an offline emissions preprocessor. 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 CESM2 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 multiscale 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 provide a general input data broker facility for CESM in the future.

1.5 CODE AND DATA AVAILABILITY

The code used in this paper is permanently archived at <https://github.com/jimmielin/HEMCO3-Paper-Code> (last access: 2 September 2021) (<https://doi.org/10.5281/zenodo.4706173>, Lin, 2021).

Input data for use with HEMCO 3.0 listed in Table 1.1 can be obtained from the GEOS-Chem data repository at <http://ftp.as.harvard.edu/gcgrid/data/ExtData/HEMCO/> (last access: 2 September 2021).

1.6 AUTHOR CONTRIBUTIONS AND ACKNOWLEDGMENTS

DJJ, SDE, and LKE provided project oversight. HL, EWL, MPS, and CAK developed the HEMCO 3.0 code. HL, EWL, MPS, and CAK developed the HEMCO to the GEOS-Chem “Classic” interface. EWL and CAK developed the HEMCO MAPL interface for use in GCHP. CAK developed the MAPL gridded component for use in GEOS. HL, TMF, 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.

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. Computing and data storage resources, including the Cheyenne supercomputer (<https://doi.org/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.

This research has been supported by the Atmospheric Chemistry Program of the US National Science Foundation (grant no. 1914903).

2

An Adaptive Auto-Reduction Solver for Speeding Up Integration of Chemical Kinetics in Atmospheric Chemistry Models: Implementation and Evaluation in the Kinetic Pre-Processor (KPP) Version

3.0.0

KINETIC INTEGRATION OF LARGE AND STIFF CHEMICAL MECHANISMS IS A COMPUTATIONAL BOTTLENECK IN MODELS OF ATMOSPHERIC CHEMISTRY. It requires implicit solution of the cou-

pled system of kinetic differential equations with time-consuming construction and inversion of the Jacobian matrix. We present here a new version of the Kinetic Pre-Processor (KPP 3.0.0) for fast integration of chemical kinetics featuring a range of improvements over previous versions in performance, diagnostics, versatility, and community openness. KPP 3.0.0 includes a new adaptive auto-reduction solver to decrease the size of any mechanism locally and on the fly under conditions where full complexity is not needed, by partitioning species as “fast” or “slow” based on their local production and loss rates. Previous implementations of this adaptive solver suffered from excessive overhead in the repeated construction of the local Jacobian matrix or were hard-wired to specific mechanisms. Here we retain the general applicability of the method to any mechanism and avoid overhead by using pre-computed Jacobian matrix terms for the full mechanism and cropping the matrix locally to remove the slow species with no change in memory allocation. We apply this adaptive solver within KPP 3.0.0 to the GEOS-Chem global 3-D model of atmospheric chemistry and demonstrate a 32% reduction in solver time while maintaining a mean error lower than 1% for key species in the troposphere.

2.1 INTRODUCTION

Modeling atmospheric chemistry is a grand computational challenge. Current global 3-D models of oxidant-aerosol chemistry use chemical mechanisms that may involve hundreds of coupled chemical species with lifetimes ranging from less than a second to many years. Chemical evolution in such a mechanism is computed by solving a large, stiff system of coupled non-linear ordinary differential equations (ODEs) expressing the chemical kinetics of individual species. Implicit solvers are required to accommodate the coexistence of short and long time constants but are computationally expensive because of the need for repeated construction and inversion of the Jacobian matrix (Brasseur & Jacob, 2017). Chemical integration often dominates the overall computational cost of

global 3-D atmospheric chemistry model simulations, even in massively parallel environments or using graphics processing units (GPUs) (Alvanos & Christoudias, 2017; Dawson et al., 2022; Eastham et al., 2018; Zhuang et al., 2020).

Considerable research has gone into devising algorithms to speed up chemistry solvers. A common strategy is to split the mechanism species by time scales in order to decrease the stiffness of the system (Djouad & Sportisse, 2002; Gong & Cho, 1993; Young & Boris, 1977), but this tends to be mechanism-specific and is difficult to apply in global models because of the wide range of conditions that may be experienced. Wholesale reduction of the mechanism, such as in the Super-Fast mechanism used in some climate models (Brown-Steiner et al., 2018), may lead to large errors (Kelp et al., 2022) and incorrect chemical responses to perturbations. Machine learning methods can in principle speed up chemical integration by orders of magnitude but have met with little success because of the large dimensionality of the problem resulting in error growth, in addition to requiring re-training and re-evaluation after even minor updates to the chemical mechanism (Keller & Evans, 2019; Kelp et al., 2020, 2022).

A promising approach for global models is to recognize that the full complexity of the mechanism is not needed everywhere. For example, reactions involving volatile organic compounds (VOCs) and their short-lived products typically account for much of the complexity but may be unimportant outside of continental boundary layers where the VOCs are emitted. Jacobson (1995) thus applied separate mechanisms in a global model for the urban boundary layer, the global troposphere, and the stratosphere. However, fixed geographical separation between domains can result in errors and inefficiencies by not accounting for the interactions at chemical boundaries between domains (Rastigejev et al., 2007) and not allowing for a continuum of chemical regimes from source regions to the remote atmosphere. Santillana et al. (2010) developed an adaptive mechanism reduction method in which the size of the mechanism is adjusted at each grid cell and time step, classifying species as fast (coupled) or slow (uncoupled) on the basis of their total production and

loss rates. However, the overhead involved in local definition of the reduced mechanism offset the computational gains. Sander et al. (2019) and Shen et al. (2020, 2022) improved the method by pre-compiling a limited ensemble of chemical sub-mechanisms and selecting the most appropriate sub-mechanism for use based on local conditions at each timestep, thus avoiding the overhead. Shen et al. achieved a 30%–50% reduction in computational cost compared to the full parent mechanism in a global simulation of the troposphere and stratosphere, but the selection of sub-mechanisms had to be customized to the parent mechanism.

Here we develop a mechanism-agnostic, ready-to-use method for adaptive auto-reduction of any chemical mechanism and implement it as an option in a new version 3.0.0 of the Kinetic Pre-Processor (KPP). Originally developed by Damian et al. (2002) and Sandu and Sander (2006), KPP is a software tool that automatically generates code to efficiently integrate chemical mechanisms. KPP takes in a set of human-readable input files describing the mechanism and generates Fortran 90, C, or MATLAB code to solve the corresponding system of ODEs using any of a suite of integration methods. KPP is used in many atmospheric chemistry models including MECCA within MESSy (Jöckel et al., 2010; Sander et al., 2019), WRF-Chem (Fast et al., 2006; Grell et al., 2005), the forward and adjoint GEOS-Chem models (Henze et al., 2007), and the adjoint for the CMAQ model (Hakami et al., 2007; Zhao et al., 2020). Our new version KPP 3.0.0 incorporates several performance and diagnostic updates over the previous version KPP 2.1 (Sandu & Sander, 2006) in addition to the adaptive solver option.

2.2 ADAPTIVE SOLVER FOR CHEMICAL KINETICS

Atmospheric chemistry models alternate chemical integration and transport calculations through operator splitting (Brasseur & Jacob, 2017). The chemistry solver is called for a time interval of length Δt referred to as the external time step, and returns a vector of updated concentrations $\mathbf{C}$

at the end of that time step to be operated on by transport. The kinetic integration of the mechanism by the chemistry solver is done over internal time steps $h \leq \Delta t$ to reach the desired accuracy. The chemistry solver integrates a system of N coupled nonlinear first-order ODEs of the form

$$\frac{dC_i}{dt} = P_i(\mathbf{C}) - L_i(\mathbf{C}) (i = 1, 2, \dots, N) \quad (1)$$

where N is the number of coupled species in the mechanism, $\mathbf{C}$ is the vector of species concentrations of dimension N , and $P_i(\mathbf{C})$ and $L_i(\mathbf{C})$ are the production and loss rates of species i that depend on the concentrations of other species in the mechanism through the reaction rate expressions.

The external time step in a global model is typically $\sim 10^3$ s, but many species in the mechanism have lifetimes ~ 1 s or shorter. An explicit solver would require internal time steps shorter than the lifetime of the shortest-lived species in order to achieve stability, but this is not computationally practical. An implicit solver is required. The simplest such solver is the first-order backward Euler method, which approximates the solution to Equation 1 over the internal time step h with

$$\mathbf{C}(t + h) = \mathbf{C}(t) + \mathbf{s}(\mathbf{C}(t + h))h \quad (2)$$

where we define the net source term $\mathbf{s} = \mathbf{P} - \mathbf{L}$ as a vector of functions $(P_i - L_i)$. Solving Equation 2 for the unknown quantity $\mathbf{C}(t + h)$ using the Newton-Raphson method requires the repeated construction and inversion of the $N \times N$ Jacobian matrix $\mathbf{J}$:

$$\mathbf{J} = \frac{\partial \mathbf{s}}{\partial \mathbf{C}} \quad (3)$$

The construction and inversion of this Jacobian matrix is computationally expensive. Higher-order solvers generally used in atmospheric chemistry models, such as Rosenbrock (Hairer & Wan-

ner, 1991; Rosenbrock, 1963; Sandu et al., 1997) or Gear (Jacobson & Turco, 1994), similarly require the repeated construction and inversion of the Jacobian over internal time steps. The Jacobian is typically ~90% sparse allowing for efficient sparse-matrix inversion methods (Sandu et al., 1997), so that the overall cost of construction and inverting the Jacobian scales as $\sim N$ rather than a higher power.

A way to reduce the dimensionality N of the problem is to split the mechanism into “fast” species for which the coupled implicit solution is necessary and “slow” species that may be solved independently over the external time step using a fast explicit method. Young and Boris (1977) and Gong and Cho (1993) separate species into fast and slow based on their lifetimes compared with the integration time step. However, the separation results in non-conservation of mass because the reaction rates are not computed consistently. This may not be of consequence in a regional model (as used in those applications) where the domain is ventilated by the boundary conditions, so that errors do not accumulate, but it is more problematic in a global model. Santillana et al. (2010) separated instead “fast” and “slow” species on the basis of their production and loss rates, with the slow species having sufficiently low rates that their non-conservation of mass would be inconsequential. This is more relevant for global models where concentrations and rates of short-lived VOCs become very small and they have negligible influence on other species outside of their source regions. Shen et al. (2020, 2022) used the same approach to partition species between fast and slow.

Here we also follow the partitioning method of Santillana et al. (2010). At the beginning of each external time step we calculate $\mathbf{P}$ and $\mathbf{L}$ and classify species i as fast if $\max(P_i, L_i) > \delta$ and as slow otherwise, where δ is a user-selected partitioning threshold. Fast species are assigned to the coupled implicit solver as a subset (sub-mechanism) of the full mechanism, while the evolution of slow species over the external time step beginning at t_0 is calculated using an explicit first-order approximation with first-order loss rate constant $k_i(t_0) = L_i(t_0)/C_i(t_0)$:

$$C_i(t_0 + \Delta t) = \frac{P_i(t_0)}{k_i(t_0)} + (C_i(t_0) - \frac{P_i(t_0)}{k_i(t_0)})e^{-k_i(t_0)\Delta t} \quad (4)$$

In the Santillana et al. (2010) implementation, the Jacobian had to be reconstructed locally at every external time step for the identified subset of fast species and that incurred large overhead, canceling the benefit of the method. Here we avoid the overhead by taking advantage of the pre-computed Jacobian matrix terms for the full mechanism in the KPP solver to simply remove rows and columns corresponding to the slow species. This is explained in Section 3.2 and is the key new development to make the method computationally practical.

The partitioning threshold δ is set prior to integration and is tuned to balance performance and accuracy. Previous work (Santillana et al., 2010; Shen et al., 2020, 2022) considered that in a typical tropospheric chemistry mechanism, much of the coupling is associated directly or indirectly with cycling of the hydroxyl radical (OH). OH has a daytime concentration of $\sim 10^6$ molecules cm^{-3} and a lifetime ~ 1 s, so its production and loss rates are $\sim 10^6$ molecules $cm^{-3}s^{-1}$. A species with production and loss rates that are several orders of magnitude smaller would not be expected to contribute significantly to the coupling. They found $\delta \sim 10^2 - 10^3$ molecules $cm^{-3}s^{-1}$ to be adequate after testing for performance and accuracy.

Here we include the option to dynamically define δ instead of specifying a uniform value over the domain, to account for rates varying with local conditions. This is done by identifying a target species which is considered central to the mechanism and scaling its production and loss rates to define a local partitioning threshold:

$$\delta = \alpha_{target} \max(P_{target}, L_{target}) \quad (5)$$

where P_{target} and L_{target} are the local production and loss rates of the target species, and $\alpha_{target} \ll 1$ is a user-selected coefficient that depends on the target species but is otherwise fixed for the model

domain. For example, a model may use OH as target species for daytime, and NO_2 or the nitrate radical (NO_3) for nighttime. When using OH as a target species and with $\max(P_{\text{OH}}, L_{\text{OH}}) \sim 10^6$ molecules $\text{cm}^{-3} \text{s}^{-1}$, a value $\alpha_{\text{OH}} \sim 10^{-4} - 10^{-3}$ would correspond to the criteria for δ used by Santillana et al. (2010). But $\max(P_{\text{OH}}, L_{\text{OH}})$ can in fact vary over orders of magnitude depending on pressure, UV flux, and other factors, and our local specification of δ accounts for this variability. We find that this specification of dynamic threshold improves accuracy with no significant overhead since P_{target} and L_{target} are computed at the beginning of each external time step in any case.

We also include two new options in the algorithm. First is to force individual species to remain in the coupled implicit solver even if $\max(P_i, L_i) \ll \delta$. As we will see, this may be helpful for inorganic halogen species that cycle between radical and non-radical forms across sunrise/sunset. Second is to include an “append” functionality in the algorithm so that a species initially diagnosed as slow at the beginning of the external time step can be transferred into the coupled implicit solver if it becomes fast over the course of the integration. This increases accuracy with minimum overhead.

2.3 KINETIC PRE-PROCESSOR (KPP) VERSION 3.0.0

2.3.1 KPP 3.0.0 OVERVIEW

KPP (Damian et al., 2002; Sandu & Sander, 2006) generates code for solving the chemical kinetics for a given chemical mechanism defined by a list of species, reactions, and rate constants. It is designed for speed by exploiting sparse matrix algebra and pre-computation of terms. We have made several improvements to KPP relative to the previous version 2.1 (Sandu & Sander, 2006), including the adaptive solver option. We present these improvements as KPP version 3.0.0, available for download from <https://github.com/KineticPreProcessor/KPP/> (<https://doi.org/10.5281/zenodo.7308373>) with detailed documentation at <https://kpp.readthedocs.io>. Existing features in the previous versions of KPP are maintained in KPP version 3.0.0. Our new version has zero numerical

<pre> #INTEGRATOR rosenbrock_autoreduce #LANGUAGE Fortran90 #MINVERSION 3.0.0 #AUTOREDUCE on #FAMILIES POx : O3 + NO2 + HNO3; LOx : O3 + NO2 + HNO3; #DEFVAR O3 = IGNORE; NO = IGNORE; ... #EQUATIONS O3 + hv = 2OH + O2 : PHOTOL(1); NO + O3 = NO2 + O2 : ARR(1.8d-12, -1370.0); CO + OH = CO2 + HO2 : 2.4d-13; HO2 + NO = OH + NO2 : ARR(3.7d-12, 240.0); NO2 + hv = NO + O3 : PHOTOL(2); HO2 + HO2 = H2O2 + O2 : ARR(1.2d-13, 749.0); H2O2 + hv = 2OH : PHOTOL(3); CH4 + OH = CO + HO2 : ARR(3.1d-12, -187.0); ... </pre>	} KPP Options Integrator, target language, and other code generation options Production and Loss Families (Optional) Definition of species "IGNORE" if mass balance checks are not required Otherwise specify stoichiometric composition } List of reactions Reactants = Products : Rate constant Fortran 90 expressions are supported in rate constants
---	--

Figure 2.1: Example KPP input file. The KPP input file includes options for code generation, adaptive auto-reduction, diagnostics for production and loss rates of chemical families, and a list of species and reactions. Reaction rate constants can be specified as expressions in the target language (in this case Fortran 90), here showing calls to functions calculating photolysis and Arrhenius rate expressions.

differences compared to KPP version 2.1 when using the same configuration as verified using the “saprc2006.kpp” example.

KPP takes in as input one or more text files, with an example shown in Figure 2.1. The input is not necessarily in one single file; it may be split into several files for readability. The files describe the chemical mechanism, the choice of numerical solver (e.g., Rosenbrock), target language (i.e., Fortran 90, C, or MATLAB), floating point type (single- or double-precision), production and loss diagnostics for selected chemical families (optional), and whether to use the adaptive solver option (optional). Some inputs are not part of the KPP code generation and are instead left for users to adjust at runtime, including convergence criteria (absolute and relative tolerance), numerical order of the solver, and adaptive solver options, in order to enable the user to experiment with different thresholds for performance and accuracy.

Based on the specifications in the KPP input file(s), KPP creates files in the target language con-

taining a description of the system of ODEs (number of species, reactions, and a numbered list of species), code to calculate the time derivatives, Jacobian matrix, and solution by back-substitution, along with a copy of the numerical solver (such as Rosenbrock) and supporting routines for sparse linear algebra. This set of files can be either run standalone as a box model or can be included in a 3-D model to update concentrations locally over external time steps.

Auto-reduction of the mechanism with the adaptive solver described in Section 2 is an option in KPP 3.0.0 using the Fortran 90 language, enabled in the configuration file by `#AUTOREDUCE on` and specifying the corresponding integrator (e.g., `#INTEGRATOR rosenbrock_autoreduce`). This sets up the capability for the user to reduce the mechanism locally through specification of the partitioning threshold δ between fast and slow species, listing any species for which this partitioning should not be applied. These specifications are done at runtime for flexibility. Even when the auto-reduction solver is used, mechanism auto-reduction can be disabled at runtime, defaulting to the behavior of the original integrator, to check for accuracy. A test case box model for auto-reduction is included as part of KPP 3.0.0 documentation.

In addition to the option for adaptive mechanism auto-reduction, several additional features and improvements were added to KPP 3.0.0 relative to the previous version 2.1 (Sandu & Sander, 2006). These include:

1. *Redeployment of KPP source code, continuous integration, and documentation for community development.* KPP 3.0.0 source code has been redeployed on GitHub for community development. The GitHub repository incorporates continuous integration (CI) tests which automatically compile the KPP source code to build and run combinations of sample chemistry mechanisms and integrators into box models at every code revision. This helps to ensure that new features and updates added to KPP do not break existing functionality and that numerical results for existing configurations are not affected. The documentation has also been

relocated to <https://kpp.readthedocs.io> where it is automatically built with each code revision on GitHub.

2. *New diagnostics.* The following diagnostic features have been added to KPP 3.0.0:
 - *Production and loss rates for chemical families.* KPP 3.0.0 allows for the definition of families of chemical species for computing production and loss for that family, ignoring interconversion reactions within the family. This is useful in models for example, to keep track of odd oxygen (Bates & Jacob, 2020) or nitrogen oxides. Families are defined in the #FAMILIES section of the KPP input file.
 - *Stoichiometric numbers.* The stoichiometric numbers of all reactions in the mechanism are now available in the CalcStoichNum subroutine in the KPP-generated code. This feature is used to calculate the importance of chemical species in a mechanism for the skeletal mechanism reduction in Sander et al. (2019).
 - *Individual reaction rates and time derivatives.* The reaction rates and their time derivatives are now available as optional outputs from the KPP-generated code.
3. *Addition of new solvers.* Several solvers have been added as options for integrating chemical kinetics including VODE (Brown et al., 1989), SDIRK (Hairer & Wanner, 1991), 3-stage Runge-Kutta, and forward and backward Euler methods.
4. *Addition of new rate law functions.* Rate law functions for three-body reactions using the formulas proposed by JPL (<https://jpldataeval.jpl.nasa.gov>) and IUPAC (<https://iupac.aeris-data.fr/>) have been added to the built-in rate laws in KPP 3.0.0. Rate law functions are not limited to those built in KPP and can be added by including extra source code files in the KPP input.

5. *Miscellaneous performance improvements in Fortran 90.* KPP 3.0.0 optimizes for Fortran 90 performance by applying several guidelines in coding, particularly in reaction rate calculations that are computed repeatedly in loops:

- *Unifying number precision.* Previous inputs to KPP used both single- and double-precision numbers. KPP input files and code now do not mix number precision to avoid conversion, which loses precision and costs computational time.
- *Switch to control update of reaction rate constants.* Previously, KPP called subroutines to update reaction rate constants at every internal time step, but this is computationally expensive and may not be needed in most model applications where the variables affecting rate constants are not updated between internal time steps. An optional switch has been added so that rate constants are updated only at the beginning of the external time step.
- *Optimized rate law functions.* Rate law functions have been split to avoid computation of unnecessary terms. For example, previously a single function with Arrhenius temperature dependence was used for all reactions: $ARR_{abc}(A, B, C) = A * \exp(B/T) * \exp(300/T)^C$. Many reactions may have $B = 0$ or $C = 0$ but the expressions $300/T$ or $\exp(0)$ are still needlessly computed. Separate rate functions such as $ARR_{ab}(A, B) = A * \exp(B/T)$ have been added, leading to a 44% performance improvement of reaction rate computations in a full-chemistry GEOS-Chem model run.
- *Avoiding conditionals and optional arguments.* Conditional clauses such as IF, ELSE, and SELECT CASE, and testing for optional arguments add significant computational cost if called thousands of times. If a conditional clause or optional argument is present in a frequently called subroutine, the subroutine is split into different func-

tions for each case.

- *Thread safety for generated code.* The code generated is now thread safe, so that calls for updating rate constants and running the integrator can be placed in an OpenMP parallel loop for parallelization.
- *Improved expressions for vector and array functions.* Several functions for basic vector and array operations originally used reference BLAS (Basic Linear Algebra Subprograms) implementations. These have been replaced with Fortran 90 expressions to allow for compilers to better optimize the code.

6. *Miscellaneous code improvements in KPP.* The C source code of KPP has been improved so that no compiler warnings are generated. A more consistent memory allocation helps to avoid buffer overflow problems. The KPP language is now parsed by `bison` instead of `yacc`.

2.3.2 ADAPTIVE SOLVER IMPLEMENTATION IN KPP 3.0.0

We implemented the adaptive solver as an option within KPP's Fortran 90 version of the Rosenbrock solver using sparse matrix algebra. KPP is computationally efficient because the functions to compute the time derivatives for each species and the Jacobian matrix (expressed in terms of reaction rates and species concentrations), along with all sparse matrix algebra routines, are pre-generated and use fixed indices to access species vectors and matrices. This means that the problem size and memory space is fixed at compile time, avoiding expensive memory allocation operations. However, this also yields a fixed problem structure that is difficult to manipulate. One major source of overhead in locally defined sub-mechanisms as in Santillana et al. (2010) is associated with repeatedly re-allocating and de-allocating memory to accommodate changing problem sizes for each sub-mechanism (Shen et al., 2020). Our adaptive solver implementation in KPP 3.0.0 uses a mapping operation to project the full mechanism into sub-mechanisms, thus reusing the same memory space

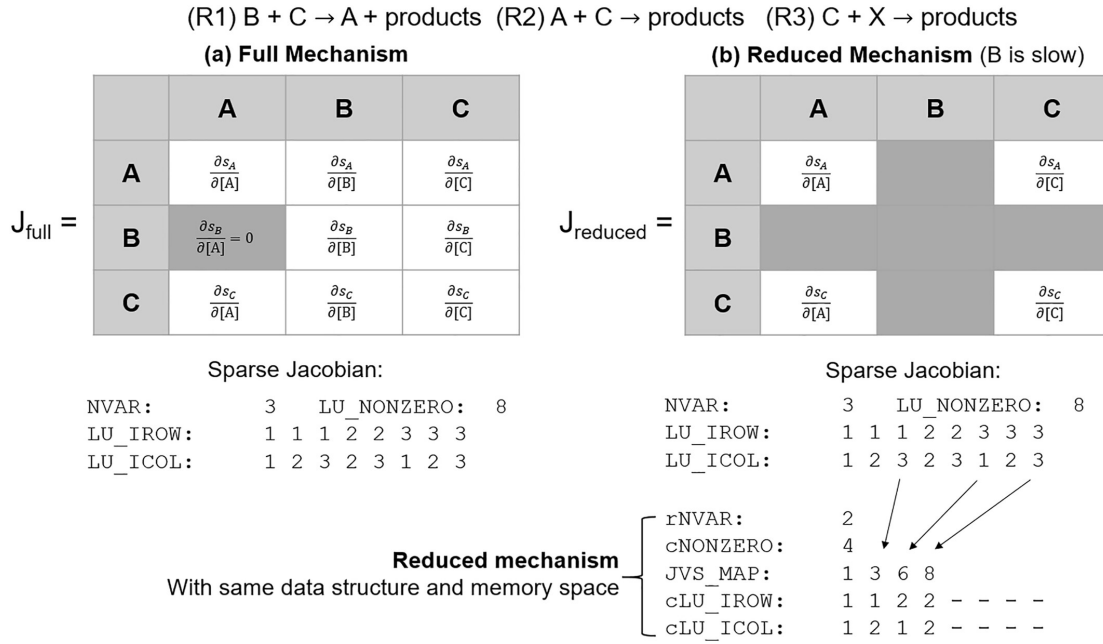


Figure 2.2: Mechanism auto-reduction in KPP. Panel (a) shows the data structure of the sparse Jacobian data within KPP for a sample 3-species mechanism. If species B is diagnosed as slow, then the corresponding rows and columns of the Jacobian are no longer calculated (panel (b)), and the indices pointing to the sparse Jacobian data (LU_IROW, LU_ICOL) are adjusted to remove the slow species (cLU_IROW, cLU_ICOL) through a mapping array (JVS_MAP) while preserving the sparse matrix in row-compressed form. The result is a reduced mechanism with the same data structure as the original one, but with smaller dimensions.

to avoid expensive resizing operations.

Figure 2.2a shows the sparse Jacobian data as stored within KPP. In this example mechanism, there are three species (NVAR) and eight non-zero entries in the Jacobian (LU_NONZERO) of the full mechanism. The row and column indices of these eight non-zero entries in Jacobian matrix are correspondingly specified in LU_IROW, LU_ICOL in row-compressed form. At the beginning of every external time step, the production and loss rates of each species are calculated and compared to the partitioning threshold to separate species into fast and slow. Figure 2.2b shows an example where species B is slow. The entries in the Jacobian corresponding to B no longer need to be computed, and a mapping operation is performed: JVS_MAP corresponds to the non-zero Jacobian matrix entries still present in the sub-mechanism consisting of fast species. The smaller sub-mechanism

can now be described by two species (`rNVAR`) and four non-zero entries in the reduced Jacobian (`cNZERO`) with indices described by `CLU_IROW` and `CLU_ICOL`. The data structure of the smaller sub-mechanism is identical to the full mechanism, and the same routines are used to solve it, without the need to generate extra code, or resizing memory.

The mechanism auto-reduction is performed once at the beginning of every external time step. The set of fast and slow species are established according to the runtime options for the partitioning threshold δ and the list of species to be excluded from partitioning and kept in the fast subset under all conditions (`keepSpcActive`). Based on the list of species in the fast set, the mapping (`JVS_MAP`) from the full mechanism to the fast sub-mechanism is created. Because KPP generates hard-coded source code to compute each term of the Jacobian matrix and back-substitution for computational efficiency, two logical control vectors, `DO_JVS` and `DO_SLV`, are created to skip computation of terms corresponding to slow species as these are no longer necessary.

The separation between fast and slow species is controlled by the initial conditions at the beginning of the external time step, but an optional “append” function is added to account for an initially slow species becoming fast over the course of the internal time steps. This function appends new species to the fast sub-mechanism and adjusts the logical control vectors if these species are initially partitioned as slow but their production or loss rate exceed the partitioning threshold over the course of the internal time stepping. Diagnosing this has little overhead, because the production and loss rates of all species are already computed at every internal time step.

We used a box model integration of the GEOS-Chem chemical mechanism (described below) to analyze the overhead of the adaptive solver implemented within KPP. By forcing the adaptive solver routines to run with a threshold of 0, we determined that the KPP overhead added by the auto-reduction is 10%–16%. The main source of overhead is the copying of data between the full and reduced sub-mechanism (Figure 2.2), where the worst-case scenario is when all species are partitioned as fast and all species’ data need to be copied between the full to the “reduced” data struc-

tures. Profiling tests show that other steps such as the partitioning of species between fast and slow, or the first-order approximation for slow species, contribute negligible overhead.

We also used the box model to verify that our adaptive solver introduces minimal mass non-conservation compared to the full solver. Integrating the box model over four sets of initial conditions from the GEOS-Chem model over 10,000 time steps shows that the total masses of odd oxygen, reactive nitrogen, chlorine, and bromine (defined in Table S1 in Supporting Information S1) are conserved within 0.30% in the standard configuration of the adaptive solver as compared with the full mechanism solver. Figure S2 in Supporting Information S1 shows the geographical distribution of mass non-conservation for these four families at a snapshot in time in a global GEOS-Chem model simulation, again comparing the adaptive solver in its standard configuration to the full solver. Mass conservation is within 0.01% for odd oxygen and reactive nitrogen and at worst within 0.2% for bromine.

2.4 ADAPTIVE MECHANISM AUTO-REDUCTION IN THE GEOS-CHEM MODEL USING KPP VERSION 3.0.0

We demonstrate the adaptive solver capability in KPP version 3.0.0 with the GEOS-Chem global 3-D atmospheric chemistry model version 13.4.0 (Bey et al., 2001; <https://doi.org/10.5281/zenodo.6511970>). The chemical mechanism includes comprehensive oxidant-aerosol chemistry in the troposphere and the stratosphere with 291 chemical species and 913 reactions. Recent updates to the mechanism include Cl-Br-I tropospheric halogen chemistry (Wang et al., 2021), isoprene chemistry (Bates & Jacob, 2019), aromatic chemistry (Bates et al., 2021), hydroxymethanesulfonate cloud chemistry (Moch et al., 2020), and NO_y cloud and aerosol chemistry (Holmes et al., 2019). Heterogeneous sulfur chemistry that was previously simulated with a separate module was brought into KPP in GEOS-Chem version 13.4.0 with the addition of new rate functions.

GEOS-Chem has been structured to interact with the KPP-generated solver code through the FlexChem interface which prepares data for the solver code, executes the code, and retrieves concentrations at the end of the time step. FlexChem allows GEOS-Chem to use modules outside of KPP for computing reaction rates, for example, interfacing with Fast-JX for photolysis rates (Bian & Prather, 2002; Mao et al., 2010). FlexChem also includes a derived type object, `State_Het`, which passes state variables from GEOS-Chem model for calculating heterogeneous chemistry reaction rates including cloud liquid water content, aerosol size distribution, pH, and alkalinity. This derived type object holds common intermediate quantities necessary for heterogeneous reaction rate computations, such as aerosol area, avoiding repeated computation and memory use. FlexChem also allows GEOS-Chem users to modify the chemical mechanism input into KPP without modifying the GEOS-Chem source code.

We evaluate the accuracy and computational performance of the adaptive solver in KPP version 3.0.0 using global GEOS-Chem simulations at $2^\circ \times 2.5^\circ$ resolution with 72 vertical levels extending up to 0.1 hPa. The model is driven by the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) meteorological fields from the NASA Global Modeling and Assimilation Office (GMAO). The external time step for the chemistry solver is 20 min. All simulations are conducted on the same single-node hardware with 24 Intel Cascade Lake physical cores (Intel(R) Xeon (TM) Platinum 8268 CPUs with a base clock speed of 2.90 GHz, no hyper-threading logical cores), 100 GB of RAM, and a high-performance Luster parallel file system. The model was compiled using Intel(R) Fortran Compiler (ifort) version 2021.2.0.

We select as the standard configuration of the adaptive solver within the GEOS-Chem model a dynamically defined threshold with OH as target species during daytime, and NO₂ during nighttime, with coefficients $\alpha_{OH} = 5 \times 10^{-5}$ and $\alpha_{NO_2} = 1 \times 10^{-4}$. We find that using NO₃ as an alternative nighttime target species results in higher accuracy but lower speed-up (Table S3 in Supporting Information S1). We do not force any species to remain in the implicit KPP solver as fast

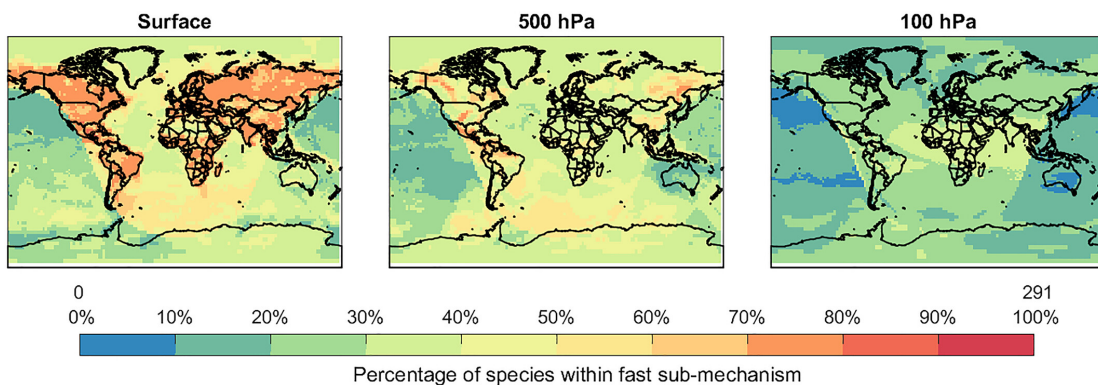


Figure 2.3: Percentage of GEOS-Chem species retained in the fast sub-mechanism when using the adaptive solver. The full GEOS-Chem mechanism has 291 species to describe tropospheric and stratospheric oxidant-aerosol chemistry. Only a fraction of species is retained as fast in the KPP solver, while the other species are solved individually using Equation 4. Results are shown for a snapshot in time on 1 July 2014, 12:00 UTC at different altitudes. The adaptive solver uses a dynamically defined threshold (Equation 5) with target species OH in daytime ($\alpha_{OH} = 5 \times 10^{-5}$) and NO₂ at night ($\alpha_{NO_2} = 1 \times 10^{-4}$).

and we do not use the append functionality. We find that we achieve a net 32% reduction in integration time in this standard configuration.

Figure 2.3 shows the percentage of species partitioned as fast (and hence retained in the KPP integration) using the adaptive solver’s standard configuration. Starting from the ensemble of 291 species in the GEOS-Chem full mechanism, we find that over 60% are partitioned as fast in surface air over land, reflecting VOC sources, whereas only 10%–50% are partitioned as fast over the ocean. The fraction of retained species decreases with altitude, and fewer than 40% are partitioned as fast in the stratosphere where VOC chemistry is mainly limited to methane. Fewer species are partitioned as fast at night than in daytime. These results are consistent with Shen et al. (2020).

We evaluate the accuracy of the adaptive solver (AS) relative to the full mechanism solver over the global GEOS-Chem domain using the relative root mean squared (RRMS) error metric. For each species i , the RRMS error is:

$$RRMS_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{j=1} \left(\frac{C_{i,j,full} - C_{i,j,AS}}{C_{i,j,full}} \right)^2} \quad (6)$$

where $C_{i,j,full}$, $C_{i,j,AS}$ are the concentrations of species i in grid box j for simulations without and with the adaptive solver. The RRMS is computed over the ensemble N_i of ordered grid boxes that account for 99% of the total mass of species i in the boundary layer (surface to PBL height from MERRA2), free troposphere (boundary layer height to tropopause), and stratosphere, respectively, and where $C_{i,j,full}$ is greater than 10 molecules cm^{-3} .

Figure 2.4a shows the mean errors at the end of a 1-year simulation for species in different chemical categories (Table S4 in Supporting Information S1) and Figures 4b–4d shows the time evolution of errors over the 1-year simulation for all species in the standard configuration of the adaptive solver within the GEOS-Chem model, respectively. The mean error for most categories is below 1% within the boundary layer and the free troposphere. There is no error growth in the troposphere but there is some in the stratosphere. Figure 2.5 shows the geographical distribution of relative errors (absolute errors shown in Figure S6 in Supporting Information S1) for ozone, OH, and NO_2 at the surface and 500 hPa for the end of the 1-year simulation. Errors are generally lower than 1% except for OH at high latitudes and NO_2 over the Southern Ocean, where errors are up to 3% for OH and 6 ~ 10% for NO_2 . Errors over land for these species are minimal.

The largest errors in Figure 2.4 are found for inorganic halogen radicals and their reservoirs. Halogen radicals cycle rapidly during the day but become locked in reservoirs at sunset, to be released again at sunrise (Wang et al., 2021). Partitioning between fast and slow species at sunrise/sunset leads to large errors and mass balance issues associated with the use of the first-order approximation (Equation 4). These problems were previously identified by Shen et al. (2020, 2022). Long residence times in the stratosphere compound the problem and drive slow error growth for other species. The impact is much less in the troposphere where the inorganic halogens are removed by

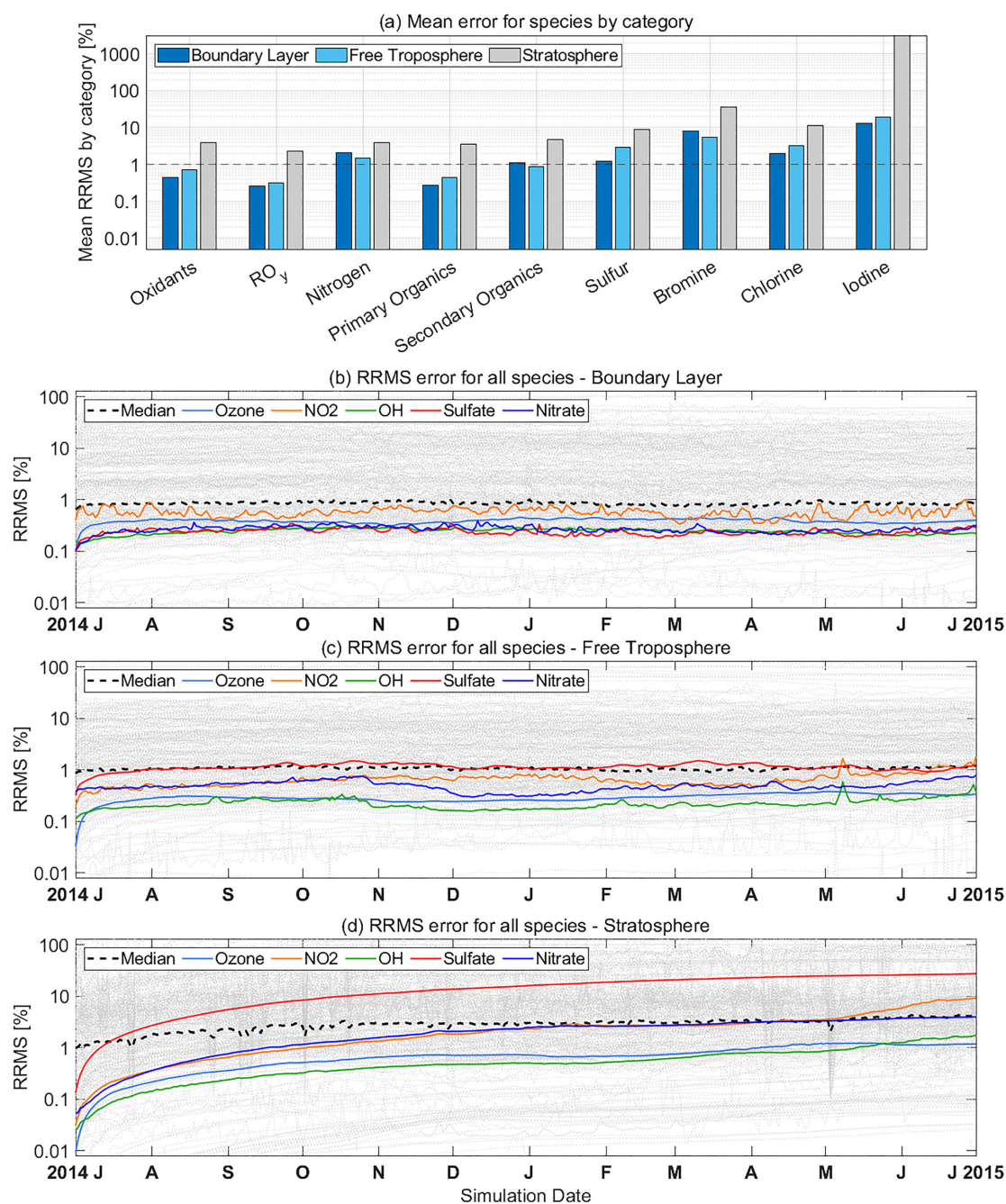


Figure 2.4: Accuracy of the adaptive solver in a 1-year GEOS-Chem simulation. Panel (a) shows the mean RRMS errors for species in different categories at the end of the 1-year simulation starting on 1 July 2014. The categories are as defined in the standard GEOS-Chem benchmarking output diagnostics (Table S4 in Supporting Information S1). Panels (b–d) show the time evolution of RRMS errors for all species in the boundary layer, free troposphere, and stratosphere, respectively, with colored lines for species of particular interest.

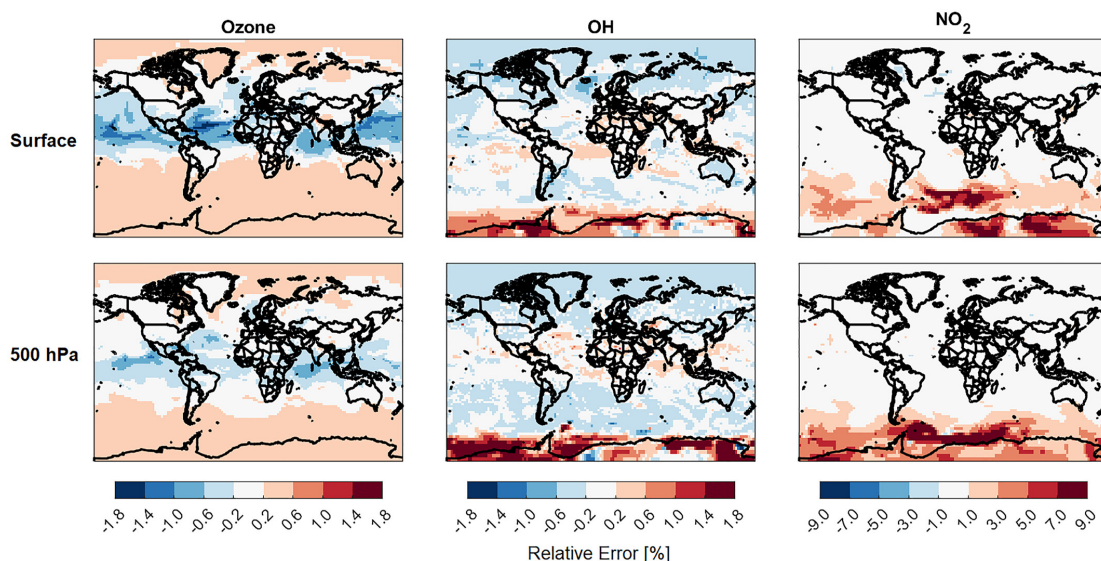


Figure 2.5: Accuracy of the adaptive solver for ozone, OH, and NO₂ global distributions. The figure shows the percent errors relative to the full solver for the daily mean concentrations on 1 July 2015, the last day of the 1-year simulation, in surface air and at 500 hPa. The absolute error values corresponding to this figure are shown in Figure S6 in Supporting Information S1.

wet deposition.

Our results in Figure 2.4 indicate that halogen species should be kept in the implicit solver as fast species if the primary interest of the simulation is the stratosphere. Figure S5 in Supporting Information S1 shows the mean errors and time evolution of errors over the 1-year simulation when halogen species are kept as fast in the adaptive solver. In this configuration, the performance improvement of the adaptive mechanism is 23%, as compared to 32% in the standard configuration. Keeping the halogen radicals and their reservoir species as fast prevents large errors from developing in the stratosphere, in addition to avoiding error growth and spikes in errors as shown in Figure 2.4d. A similar problem though not of the same magnitude is found for hydrogen oxide radicals controlling the concentration of OH in the stratosphere, and this causes high errors in stratospheric sulfate exceeding 10% above 85 hPa. The list of species to be kept in the implicit solver in the GEOS-Chem model can be configured in the FlexChem interface and passed to KPP 3.0.0 as a runtime

option. An alternative to customizing the list of species to be kept in the implicit solver is to disable the adaptive solver in the stratosphere while retaining it in the troposphere. In this case, the performance improvement from using the adaptive mechanism decreases from 32% to 20%. Such ad hoc adjustments to the fast/slow species partitioning would be mechanism-specific and can be experimented with by users. We find in GEOS-Chem that they are not needed for standard applications focusing on tropospheric chemistry.

We find in this application that the append functionality (allowing species to switch from slow to fast over internal time steps) does not provide significant error reduction, in particular for the halogen species, and degrades the performance improvement of the adaptive mechanism to 24% instead of 32%. We retain it as an option in the adaptive solver code as it may be helpful for longer external time steps (here 20 min) or other chemical mechanisms.

The adaptive solver uses a dynamically defined threshold (Equation 5) with target species OH in daytime ($\alpha_{OH} = 5 \times 10^{-5}$) and NO₂ at night ($\alpha_{NO_2} = 1 \times 10^{-4}$), and does not force any species below that threshold to remain as fast.

Our analysis of the performance and accuracy of the adaptive solver is based on 20-min external time steps for the chemistry solver in the GEOS-Chem standard $2^\circ \times 2.5^\circ$ configuration, but regional models with finer grid resolution would use finer external time steps for operator splitting with transport (Philip et al., 2016). We tested the sensitivity of our results to the size of the external time step within the GEOS-Chem $2^\circ \times 2.5^\circ$ environment as shown in Table S3 in Supporting Information S1. We find that finer external time steps result in higher accuracy but lower speed-up, which makes sense in terms of the overhead required at the beginning of each external time step to define the sub-mechanism and cull the Jacobian, but also the opportunity to update the separation between fast and slow species more frequently for better accuracy. Ultimately, this sensitivity to external time-stepping will need to be tested by users in their own model environment (and their own mechanism).

2.5 CONCLUSIONS

We presented an updated version of the (KPP 3.0.0) to integrate stiff chemical mechanism kinetics typical of atmospheric chemistry models. KPP was originally designed for flexibility and speed. KPP 3.0.0 features several improvements for performance, diagnostics, choice of solvers, and code openness. It includes an adaptive solver capability for mechanism auto-reduction where and when the full mechanism is not needed.

The adaptive solver performs auto-reduction of the chemical mechanism locally and on-the-fly at runtime, by comparing the local production and loss rates of each species with a partitioning threshold (δ). Species with production and loss rates higher than the threshold are considered fast and are solved as a coupled sub-mechanism within KPP, while other species are considered slow and solved individually by an explicit method. Previous application of this adaptive solver method suffered from large overhead due to the need for local reconstruction of the reduced Jacobian matrix and the associated memory allocation and deallocation. We solved this problem here by using pre-computed Jacobian terms for the full mechanism with a mapping operation to crop rows and columns corresponding to the slow species without changing the memory allocation.

KPP 3.0.0 features additional improvements for performance, diagnostics, versatility, and openness. Improved performance includes more efficient calculation of reaction rates from the KPP rate functions and thread safety for parallelization. New diagnostics include individual reaction rates, production and loss rates for chemical families, and stoichiometric numbers. Improved versatility includes expanded choice of chemical solvers. KPP 3.0 is now hosted on GitHub (<https://github.com/KineticPreProcessor/KPP>) to enable community access, development, and testing.

We evaluated the adaptive solver implemented in KPP 3.0.0 by conducting a 1-year simulation with the global 3-D GEOS-Chem atmospheric chemistry model including 291 species in the full mechanism for oxidant-aerosol chemistry in the troposphere and stratosphere. Results show that a

32% performance improvement in the solver can be achieved with a target error of 1% for key species in the troposphere. Errors in the stratosphere can be larger, driven by halogen chemistry. Lower errors especially in halogen species can be achieved by keeping these species within the fast sub-mechanism but this reduces the performance improvement to 23% and is mainly beneficial in the stratosphere.

The release of KPP 3.0.0 introduces improvements in development infrastructure, diagnostics, and performance, particularly in Fortran 90 applications. However, one of the strengths of the KPP software is the capability to generate code for different programming languages. Development directions for future versions include (a) adding support for modern languages such as Python and Julia; (b) refactoring of the generated code to avoid global data structures for easier parallelization; (c) streamlining inputs and outputs of all integrators for consistency; (d) supporting the adaptive solver option in other integrators and programming languages; and (e) improving interaction and compatibility with the Master Chemical Mechanism (<http://mcm.york.ac.uk>). These improvements will allow KPP to better serve the community as a versatile tool for solving chemical kinetics within atmospheric chemistry models and other applications.

2.6 CODE AND DATA AVAILABILITY

Source code for KPP 3.0.0 is available at <https://github.com/KineticPreProcessor/KPP> (Sandu et al., 2022; <https://doi.org/10.5281/zenodo.7308373>). The adaptive solver box model is available at <https://github.com/KineticPreProcessor/KPP-AR-boxmodel/> (Long & Lin, 2022; <https://doi.org/10.5281/zenodo.6791657>). The adaptive solver implementation within the GEOS-Chem atmospheric chemistry model used in this work is available at <https://github.com/jimmielin/geos-chem/tree/staging/autoreducekpp> (The International GEOS-Chem Community, 2022; <https://doi.org/10.5281/zenodo.6791655>).

2.7 ACKNOWLEDGMENTS

This work was supported by the US EPA Science to Achieve Results (STAR) Program, by the NASA Modeling, Analysis, and Prediction (MAP) Program, and by the NASA Atmospheric Composition Modeling and Analysis Program (ACMAP).

3

Intercomparison of GEOS-Chem and CAM-chem tropospheric oxidant chemistry within the Community Earth System Model version 2

TROPOSPHERIC OZONE IS A MAJOR AIR POLLUTANT AND GREENHOUSE GAS. It is also the primary precursor of OH, the main tropospheric oxidant. Global atmospheric chemistry models show large differences in their simulations of tropospheric ozone budgets. Here we implement the widely used GEOS-Chem atmospheric chemistry module as an alternative to CAM-chem within the Community Earth System Model version 2 (CESM2). We compare the resulting simulations of tropospheric ozone and related species to observations from ozonesondes, the ATom-1 aircraft campaign

over the Pacific and Atlantic, and the KORUS-AQ aircraft campaign over the Seoul Metropolitan Area. We find that GEOS-Chem and CAM-chem within CESM2 have similar tropospheric ozone budgets and concentrations usually within 5 ppb but important differences in the underlying processes including (1) photolysis scheme (no aerosol effects in CAM-chem), (2) aerosol nitrate photolysis, (3) N_2O_5 cloud uptake, (4) tropospheric halogen chemistry, and (5) ozone deposition to the oceans. Global tropospheric OH concentrations are the same in both models but there are large regional differences reflecting the above processes. Carbon monoxide is lower in CAM-chem (and lower than observations) because of higher OH concentrations in the northern hemisphere and insufficient production from isoprene oxidation in the southern hemisphere. CESM2 does not scavenge water-soluble gases in convective updrafts leading to some upper tropospheric biases. Comparison to KORUS-AQ observations shows successful simulation of oxidants under polluted conditions in both models but suggests insufficient boundary layer mixing in CESM2. The implementation and evaluation of GEOS-Chem in CESM2 contributes to the MUSICA vision of modularizing tropospheric chemistry in Earth system models.

3.1 INTRODUCTION

Ozone is a central species in atmospheric chemistry. It is a major air pollutant and greenhouse gas, and the primary source of the hydroxyl radical (OH) which is the main tropospheric oxidant (Monks et al., 2015). It is produced within the troposphere by nonlinear photooxidation processes involving hydrogen oxide radicals ($\text{HOx} \equiv \text{OH} + \text{peroxy}$), nitrogen oxide radicals ($\text{NO}_x \equiv \text{NO} + \text{NO}_2$), volatile organic compounds (VOCs), and ozone itself. It is extensively observed from surface sites, aircraft, sondes, and satellites, and is thus an important indicator of skill for chemical transport models (Hu et al., 2017). At the same time, comparisons with observations can be successful for the wrong reasons. Extensive intercomparisons of global models often show similar

tropospheric ozone burdens but large differences in chemical source and sink magnitudes (Wu et al., 2007; Young et al., 2018), implying large differences in sensitivity to perturbations. This is a particular problem for chemistry-climate models that aim to quantify chemical feedbacks on climate change.

Here we compare two state-of-science atmospheric chemistry modules, GEOS-Chem and CAM-chem, within the Community Earth System Model (CESM2) (Danabasoglu et al., 2020). CAM-chem is the resident atmospheric chemistry module in CESM2 (Lamarque et al., 2012; Tilmes et al., 2015, 2016; Emmons et al., 2020). GEOS-Chem is widely used as an offline chemical transport model (CTM) driven by external meteorological data (Bey et al., 2001). GEOS-Chem is grid-independent and modularized, so that the chemical module describing local operations in 1-D model columns (including emissions, chemistry, and deposition) is separated from the transport module (Long et al., 2015). This allows independent implementation of the GEOS-Chem chemical module in other models, including online applications where chemistry is coupled to transport (Hu et al., 2018; Lin et al., 2020; Lu et al., 2020; Keller et al., 2021). The GEOS-Chem chemical module has been previously coupled to the WRF and GEOS meteorological models to investigate aerosol-chemistry-climate feedbacks (Feng et al., 2021; Moch et al., 2022). The same scientific code base is used in the offline CTM such that version updates developed for the CTM can be seamlessly passed on to the online applications.

Fritz et al. (2022) implemented the GEOS-Chem chemical module in CESM2 as the first application of that module to an open-source Earth system model (ESM) for community use. GEOS-Chem offers an alternative representation of atmospheric chemistry to CAM-chem within CESM2, contributing to the MUSICA (Multi-Scale Infrastructure for Chemistry and Aerosols; Pfister et al., 2020) vision for CESM of allowing users to choose among a range of options for atmospheric chemistry. The GEOS-Chem emission component (HEMCO; Keller et al., 2014) has been previously implemented in MUSICA (Lin et al., 2021). Fritz et al. (2022) presented general comparisons

between GEOS-Chem and CAM-chem in the CESM2 environment. They found good agreement between the two modules for stratospheric ozone, but 10–30% lower tropospheric ozone in GEOS-Chem due to bromine chemistry not considered in CAM-chem. They found several challenges in the implementation of the GEOS-Chem chemical module within CESM2. For example, CESM2 uses the MAM4 (Modal Aerosol Model version 4; Liu et al. 2016) modal aerosol microphysics to simulate aerosol-cloud interactions and aerosol-radiation interactions, while GEOS-Chem uses either bulk or sectional representations of aerosol microphysics. CESM2 does not explicitly include scavenging of water-soluble species in convective updrafts other than through MAM4, but this is a major process in wet deposition in the GEOS-Chem CTM to prevent unphysical buildup of soluble species in the upper troposphere (Balkanski et al., 1993; Liu et al., 2001). Indeed, Fritz et al. (2022) found large overestimates of upper tropospheric aerosol in GEOS-Chem within CESM2 as compared to the offline GEOS-Chem.

Our work builds on the Fritz et al. (2022) initial implementation of GEOS-Chem in CESM2 to address the previous challenges and to give a more thorough evaluation with observations and intercomparison with CAM-chem. We focus on tropospheric ozone and related oxidant chemistry from both a global perspective (ozonesonde and ATom-1 aircraft observations) and polluted conditions over East Asia (KORUS-AQ aircraft observations). KORUS-AQ, conducted in May-June 2016, is of particular interest because of the previously identified large differences between offline GEOS-Chem and CAM-chem in simulating the aircraft observations including 20–30 ppb differences in ozone (Park et al., 2021). We analyze the individual processes driving differences between GEOS-Chem and CAM-chem, and use observations to arbitrate when possible. As part of resolving differences in photolysis rates we implement into CAM-chem the Fast-JX photolysis scheme used in GEOS-Chem (Bian and Prather, 2002), further contributing to the MUSICA vision of process-level modularization of atmospheric chemistry models.

3.2 MODEL DESCRIPTION AND METHODS

3.2.1 CESM2, CAM-CHEM, AND HEMCO

We use a beta version of CESM 2.3 including the CAM6 Community Atmosphere Model (CAM tag version `cam6_3_095`), which has provided the basis for the integration of the GEOS-Chem module into the mainline CESM code. All simulations are for the year 2016 with an 18-month initialization period. The year was chosen for evaluation with the ATom (Wofsy et al., 2018) and KORUS-AQ (Crawford et al., 2021) aircraft campaigns. We use a global grid with 32 vertical layers up to 2 hPa. The model is nudged to reanalysis meteorology (using the "FCnuded" configuration in CAM6) from 3-hourly MERRA2 data available from the NASA Global Modeling and Assimilation Office.

CAM-chem is the standard chemistry representation in CESM2 with tropospheric-stratospheric chemistry, currently using the MOZART-TS1 (Model for OZone And Related chemical Tracers; Emmons et al., 2020) mechanism and the Modal Aerosol Model with 4 modes (MAM4; Liu et al., 2016) as default. MOZART-TS1 includes 229 chemical species and 541 reactions. Photolysis is calculated using a lookup table based on the Tropospheric Ultraviolet and Visible (TUV) radiation model, which takes into account the impact of clouds but not aerosols (Kinnison et al., 2007). A sensitivity simulation developed for this project uses Fast-JX instead of the TUV look-up table for photolysis.

The CAM-chem version in our work uses HEMCO for emissions but is otherwise unmodified. HEMCO is the standard emission component of GEOS-Chem (Keller et al., 2014), now implemented in CESM as part of MUSICA (Lin et al., 2021). It allows the use of any emission inventories on any grid to be supplied to the model in netCDF format at runtime with options to add, supersede, and scale emissions. Here we use the same emissions in GEOS-Chem and CAM-chem processed through HEMCO. This includes global anthropogenic emissions from the CEDSv2 in-

ventory (Community Emissions Data System; McDuffie et al., 2021) superseded by the KORUSv5 inventory (Woo et al., 2020) over East Asia. Fire emissions are from the GFED4.1s inventory (van der Werf et al., 2017; Randerson et al., 2018). HEMCO has extensions to use emission modules dependent on environmental variables and this is applied to soil NO_x emissions from Hudman et al. (2012) and ocean iodine emissions from Sherwen et al. (2016a, 2016b). We otherwise use emissions computed from other modules in CESM to enforce consistency of the atmospheric chemistry simulation with other CESM components. This includes biogenic VOC emissions from MEGANv2.1 (Guenther et al., 2012) computed with the Community Land Model (CLM) and lightning NO_x, dust, and sea salt emissions from CAM (Price et al., 1997; Mahowald et al., 2006a, 2006b; Lamarque et al., 2012).

3.2.2 GEOS-CHEM WITHIN CESM2

Unless explicitly written otherwise, GEOS-Chem in this work refers to the online implementation of the GEOS-Chem chemical module within the CESM2 model and not the offline CTM. We use GEOS-Chem version 14.1.1 (doi:10.5281/zenodo.7696632) with the addition of particulate nitrate (pNO₃⁻) photolysis following Shah et al. (2023), which was subsequently implemented in version 14.2.0 (doi:10.5281/zenodo.8411433). The same GEOS-Chem chemical module and MERRA-2 meteorological fields are used in CESM2 and in the offline CTM simulations presented here. The GEOS-Chem chemical mechanism has 286 species and 914 reactions with a development history independent of MOZART-TS1. It features recent major updates to NO_x heterogeneous and cloud chemistry (Holmes et al., 2019), isoprene chemistry (Bates and Jacob, 2019), aromatic chemistry (Bates et al., 2021), and Cl-Br-I tropospheric halogen chemistry (Wang et al., 2021). Photolysis is calculated using the Fast-JX model (Bian and Prather, 2002) with consistent aerosol and overhead column ozone information from the GEOS-Chem simulation (Eastham et al., 2014). No aerosol microphysics is included here so that aerosol concentrations are represented by the bulk masses of

their chemical components (Park et al., 2004; Pai et al., 2020) but with four size bins for dust and two for sea salt aerosol (Alexander et al., 2005; Fairlie et al., 2010).

Fritz et al. (2022) describe the original implementation of GEOS-Chem within CESM2. They developed an interface to pass input data to GEOS-Chem, run the GEOS-Chem chemical module, and export the updated chemical species concentrations. The interface converts between the bulk aerosols in GEOS-Chem and the modal aerosols in MAM4 for aerosol-radiation and aerosol-cloud interactions. Coupling of the GEOS-Chem chemical module to CESM2 required the adaptation of several components for compatibility with CESM2 or consistency with CAM-chem. We summarize in Table 3.1 the important differences between the atmospheric chemistry representations in CAM-chem, GEOS-Chem within CESM2, and the offline GEOS-Chem CTM.

Here we make several improvements and corrections to the original implementation of GEOS-Chem within CESM2 by Fritz et al. (2022). We simulate nucleation in MAM4 by passing the gas-phase H_2SO_4 production rate computed in GEOS-Chem from the $\text{SO}_2 + \text{OH}$ reaction. We add an aerosol sink in the upper troposphere and lower stratosphere following Hodzic et al. (2015, 2016) to compensate for CESM2's omission of coupling convective transport and scavenging. We correct the sea surface temperatures passed to HEMCO, which results in inorganic iodine emissions being 1% of the previous incorrectly calculated value. We also add numerous GEOS-Chem diagnostics for analyzing model output, including individual reaction rates and total production and loss rates for individual species.

Table 3.1: Major differences between CAM-chem and GEOS-Chem simulations.

Simulation	CAM-chem within CESM2	GEOS-Chem within CESM2	Offline GEOS-Chem CTM
Meteorology	CESM2.3 nudged to MERRA2 ^a		MERRA2
Chemistry mechanism	MOZART-TS1 229 species and 541 reactions O _x -NO _x -VOC-aerosol	GEOS-Chem v14.1.1 286 species and 914 reactions O _x -NO _x -VOC-halogen-aerosol	
Photolysis	TUV lookup table	Fast-JX	
Aerosol microphysics	MAM4 modal aerosols ^b		Bulk aerosols ^c
Aerosol composition	Sulfate SOA (five VBS bins) Primary organic matter Black carbon Soil dust (three modes) Sea salt (three modes)	Sulfate, Nitrate, Ammonium SOA (four VBS bins ^d) Primary organic carbon Black carbon Soil dust (four size bins) Sea salt (two size bins)	
Dry deposition velocities (over land)	Computed by CLM		Computed by GEOS-Chem
Dry deposition velocities (over ocean and sea ice)	Computed by CAM	Computed by GEOS-Chem	
Wet deposition	Gases: Neu scheme ^e Aerosols: MAM4	Gases: Neu scheme ^e Aerosols: scavenged as HNO ₃	GEOS-Chem wet deposition scheme ^f
Scavenging in convective updrafts	Not explicitly simulated (see Section 6)		Explicitly simulated
Lightning NO _x parameterization	Price et al. (1997); 2.8-3.0 Tg N a ⁻¹		Murray et al. (2012); 5-6 Tg N a ⁻¹

^{a.} with 50-h relaxation time.

^{b.} GEOS-Chem bulk aerosols masses are mapped to MAM₄ modes for aerosol-radiation and aerosol-cloud interaction effects within CESM2. See Fritz et al. (2022) for the species mapping between GEOS-Chem species to MAM₄ aerosols.

^{c.} Sectional aerosol microphysics are available in GEOS-Chem (Yu and Luo, 2009; Kodros and Pierce, 2017) but are not used here.

^{d.} SOA ≡ secondary organic aerosol, VBS ≡ volatility basis set. GEOS-Chem here uses the Complex SOA option from Pye et al. (2010).

^{e.} Neu and Prather (2012).

^{f.} Liu et al. (2001) for water-soluble aerosols and Amos et al. (2012) for gases.

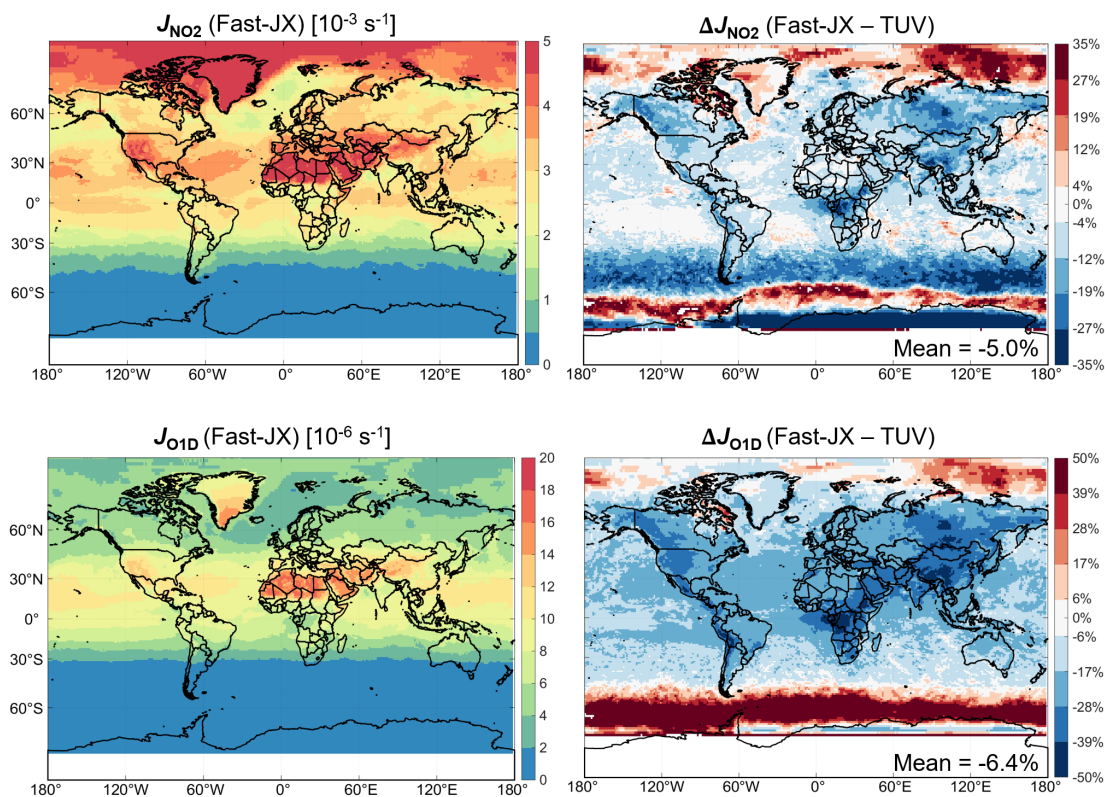


Figure 3.1: Mean photolysis frequencies for NO₂ (J_{NO_2}) and O₃ to O(¹D) (J_{O1D}) in surface air in July 2016. The left panels show the values computed by Fast-JX within GEOS-Chem. The right panels show the differences (Δ) with the values computed by TUV within CAM-chem.

3.3 COMPARISON OF PHOTOLYSIS SCHEMES: FAST-JX AND TUV

Figure 3.1 shows the mean photolysis frequencies (J values) for NO₂ (J_{NO_2}) and O₃ to O(¹D) (J_{O1D}) simulated by the GEOS-Chem model (with Fast-JX) and the difference with CAM-chem (with TUV lookup table) in surface air in July. Photolysis rates in GEOS-Chem with Fast-JX are generally lower than in CAM-chem with TUV. Differences for J_{NO_2} are typically 0–10% over oceans and 10–20% over land, while differences for J_{O1D} are typically 10–20% over oceans and 20–40% over land. There are some larger differences in polluted and open fire regions and at high latitudes.

Fast-JX and TUV use the same spectroscopic data from the NASA JPL recommendations (Burkholder

et al., 2020). Fast-JX includes aerosol extinction but TUV does not, which explains the larger differences over polluted and open fire regions. Differences over the oceans are mainly due to clouds. While Fast-JX and TUV both represent effects of cloud extinction, treatment of cloud scattering between the two schemes can be different. The effects of aerosol-cloud interactions on cloud properties through MAM₄ cause GEOS-Chem and CAM-chem to have different cloud optical depths that can lead to further differences. Cloud effects are particularly large at high latitudes because of low Sun angles. Sensitivity simulations for clear sky (no cloud or aerosol extinction input to the photolysis schemes) show smaller differences between Fast-JX and TUV, generally less than 10% for J_{NO_2} and less than 20% for $J_{\text{O}_1\text{D}}$, while sensitivity simulations with Fast-JX implemented in CAM-chem show less than 5% differences for J_{NO_2} and $J_{\text{O}_1\text{D}}$ everywhere compared to Fast-JX implemented in GEOS-Chem.

Figure 3.2 shows photolysis frequencies from the KORUS-AQ and ATom-1 campaigns derived from actinic flux measurements (Hall et al., 2018; Crawford et al., 2021), compared to the photolysis frequencies computed by Fast-JX and TUV sampled along the aircraft flight tracks. J_{NO_2} values agree within 10% and there is no systematic bias relative to observations. Fast-JX values tend to be higher than TUV at high altitudes and this can be attributed to cloud effects as discussed above. $J_{\text{O}_1\text{D}}$ values also show good agreement for ATom-1 but observed values for KORUS-AQ are much lower than for ATom-1 in the same season, which is captured by Fast-JX but not by TUV (which is 30% too high). We find that the overestimate of $J_{\text{O}_1\text{D}}$ by TUV during KORUS-AQ is due in part to not accounting for aerosol extinction. Comparison of clear-sky J values shows that there is some additional unidentified factor causing TUV to be too high during KORUS-AQ and this disappears when Fast-JX is implemented in CAM-chem. In what follows the CAM-chem simulation uses the TUV lookup table but we will comment as appropriate on the effect of switching to Fast-JX.

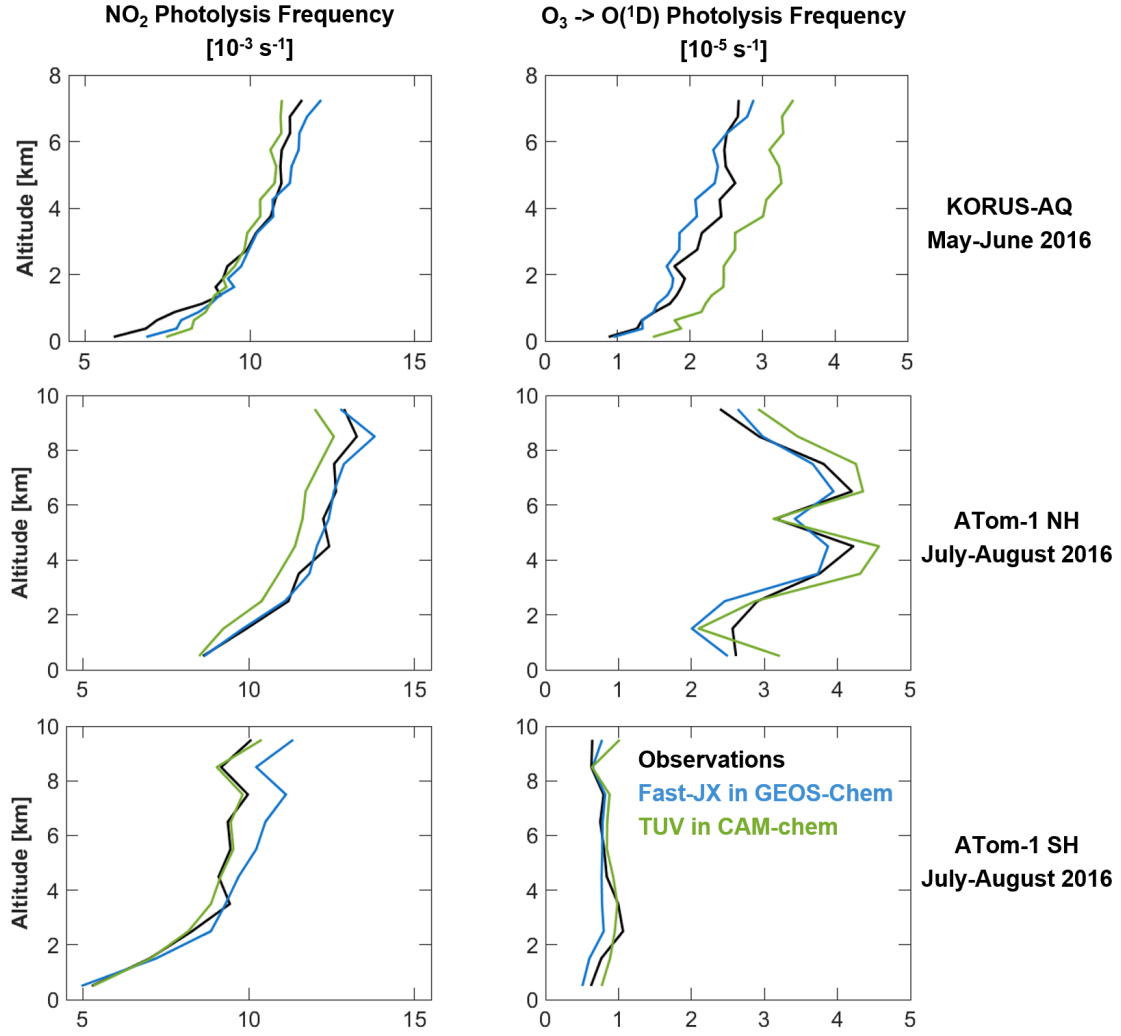


Figure 3.2: Median vertical profiles of J_{NO_2} and $J_{\text{O}^1\text{D}}$ from the KORUS-AQ aircraft campaign over the Seoul Metropolitan Area (SMA; 37–37.6°N, 126.6–127.7°E) in May-June 2016 and the ATom-1 aircraft campaign over the Pacific and Atlantic Oceans in July-August 2016. Observations from in situ measurements of actinic fluxes (Hall et al., 2018) are compared to CESM2 values using the Fast-JX scheme within GEOS-Chem and the TUV scheme within CAM-chem. The model values are sampled along the aircraft flight tracks.

3.4 GLOBAL BUDGETS AND DISTRIBUTIONS OF TROPOSPHERIC OXIDANTS

Table 3.2 shows global tropospheric ozone and OH budgets from GEOS-Chem and CAM-chem compared to the literature. Ozone budgets from the two models and the multi-model mean in Young et al. (2018) are within 10% of each other. The larger chemical production and shorter chemical lifetime in GEOS-Chem are mainly due to photolysis of particulate nitrate (Shah et al., 2023), without which chemical production in GEOS-Chem decreases by 10% to 4902 Tg a^{-1} and the tropospheric ozone burden decreases by 5% to 332 Tg . The lower dry deposition in GEOS-Chem reflects lower ozone deposition to the ocean (Pound et al., 2020). GEOS-Chem and CAM-chem have the same global OH concentrations, on the high end of the range of values from the ACCMIP and CCMi model ensembles (Naik et al., 2013; Zhao et al., 2019). The lifetime of methylchloroform against loss to tropospheric OH is 5.4 and 5.3 years respectively in GEOS-Chem and CAM-chem, 15% lower than 6.3 ± 0.4 years inferred from observations (Prather et al., 2012).

Figure 3.3 shows the spatial distribution of annual mean OH concentrations simulated by GEOS-Chem and the difference with CAM-chem. Despite having the same global mean OH concentrations, the two models have large regional differences. GEOS-Chem is up to 30% lower than CAM-chem over the continents, particularly over polluted regions, due to lower $J_{\text{O}_3\text{D}}$ and possibly higher OH reactivity. Over the Amazon and Congo basins where NO_x is low, isoprene does not titrate OH in GEOS-Chem due to recent updates in isoprene oxidation chemistry (Bates and Jacob, 2019).

Figure 3.4 shows annual mean surface and 500 hPa ozone and NO_x concentrations simulated by GEOS-Chem and differences with CAM-chem. Ozone differences are generally smaller than 5 ppb, indicating a remarkable degree of agreement. The largest surface differences are at southern mid-latitudes due to slower ozone deposition to the ocean in GEOS-Chem. At 500 hPa, GEOS-Chem has lower ozone at high latitudes due to tropospheric halogen chemistry not represented in CAM-chem, while ozone and NO_x are higher over the tropics mainly due to particulate nitrate photolysis.

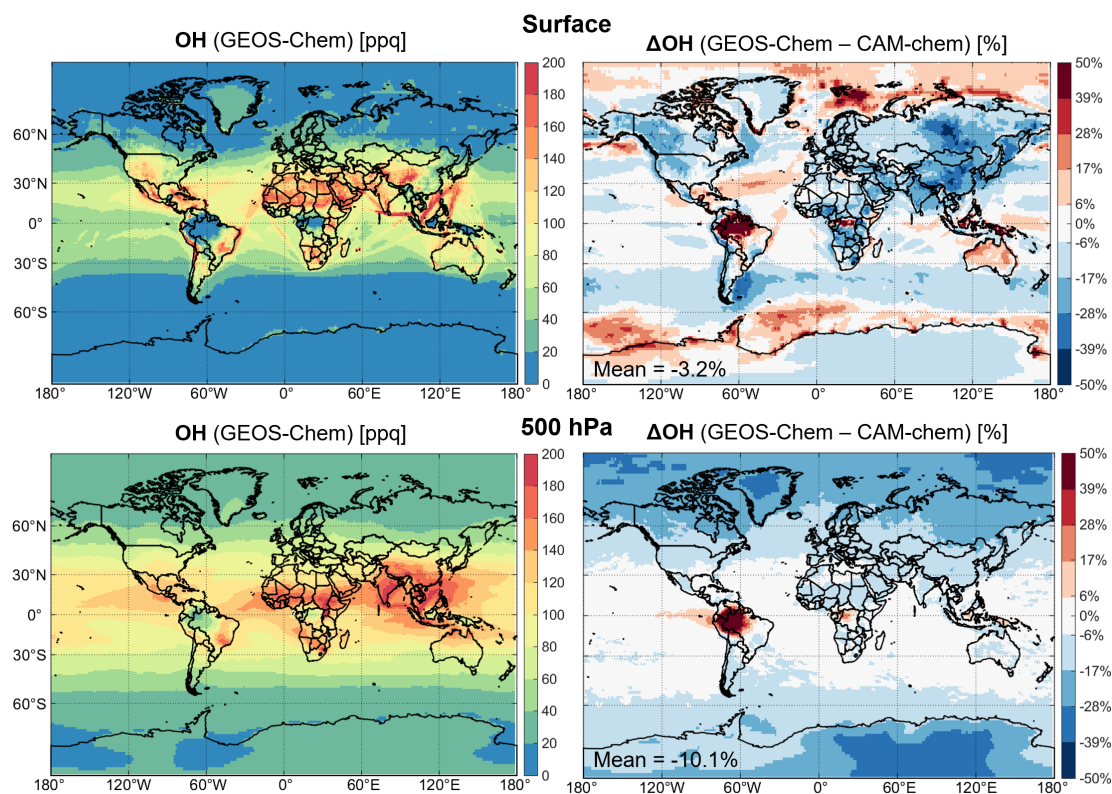


Figure 3.3: Annual mean OH concentrations in surface air and at 500 hPa in GEOS-Chem within CESM2, and differences with CAM-chem. Percentage differences are relative to CAM-chem. Global mean differences are in legend. Values are for 2016.

NO_x is lower in GEOS-Chem at high latitudes due to loss to halogen nitrates (Wang et al., 2021) and N_2O_5 uptake in clouds (Holmes et al., 2019), not included in CAM-chem.

Table 3.3 shows global budget terms in the troposphere from sensitivity simulations varying the most important differences between GEOS-Chem and CAM-chem. The largest controlling factors for tropospheric ozone differences between GEOS-Chem and CAM-chem are nitrate photolysis and tropospheric halogen chemistry, which increase and decrease the tropospheric ozone budget by 5% and 4% respectively. Particulate nitrate photolysis is necessary in GEOS-Chem to correct a low NO_x bias over the oceans in comparison to ATom data (Shah et al., 2023), but CAM-chem NO_x is 4% higher than GEOS-Chem despite not having particulate nitrate photolysis. Inspection of Table 3.3 indicates that this reflects both the lack of tropospheric halogen chemistry in CAM-chem and the use of TUV for photolysis. Using Fast-JX for photolysis in CAM-chem results in a 7% decrease in tropospheric NO_x , which we attribute tentatively to lower J_{NO_2} in surface air over continents (Figure 3.1).

Fritz et al. (2022) previously found tropospheric ozone in GEOS-Chem to be 30% lower than CAM-chem in the extratropics because of halogen chemistry, but iodine emissions in that simulation were 100-fold too high (Section 3.2.2). With corrected iodine emissions we find only a 4% decrease of tropospheric ozone in GEOS-Chem due to tropospheric halogen chemistry, which is lower than the 11–19% effect previously reported in offline GEOS-Chem simulations (Sherwen et al., 2016b; Wang et al., 2021). We find that this is due to weaker wind speeds and lower sea surface temperatures in CESM2, resulting in weaker sea salt and gaseous iodine emissions.

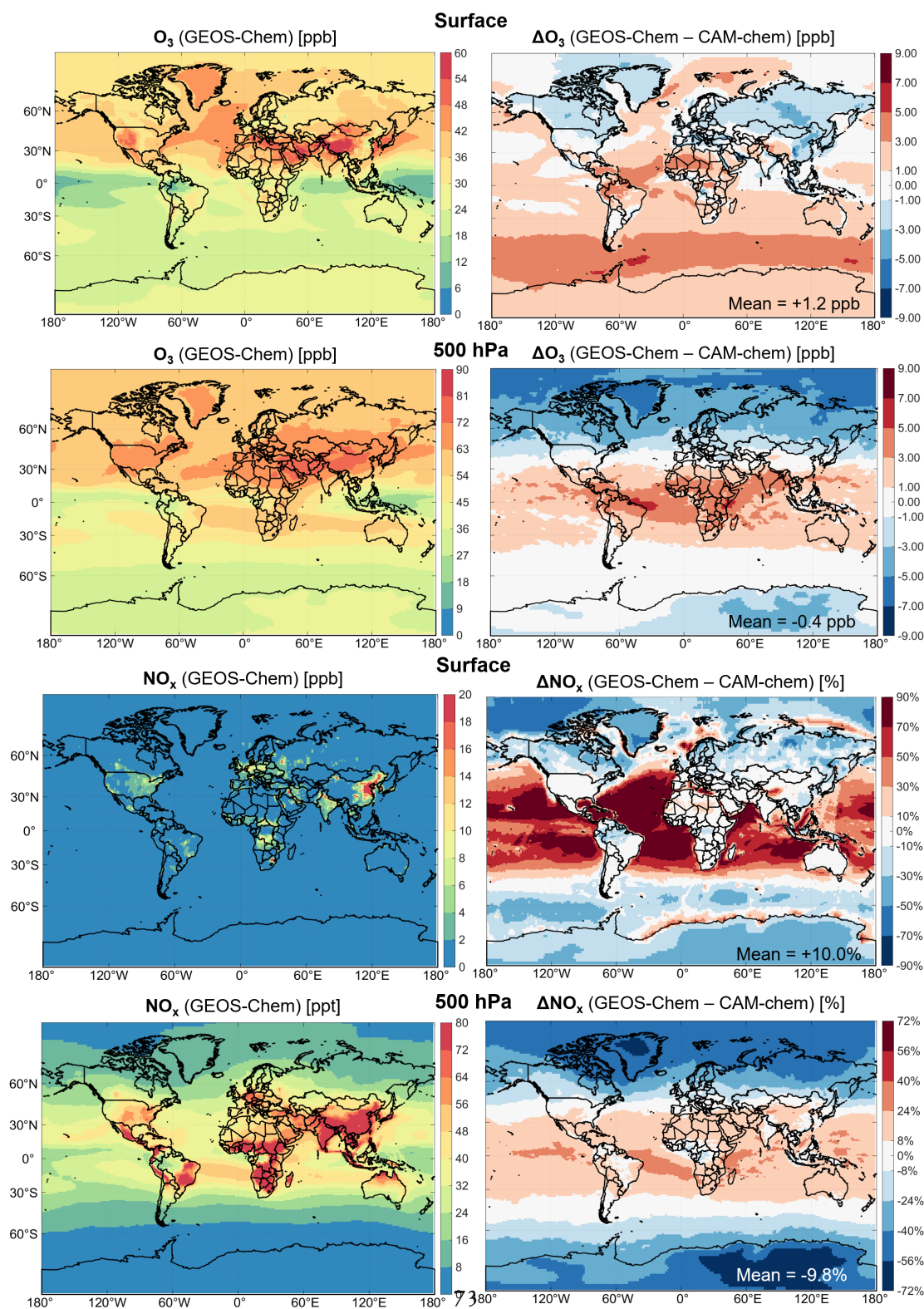


Figure 3.4: Annual mean ozone and NO_x concentrations in surface air and at 500 hPa in GEOS-Chem within CESM2, and differences with CAM-chem. Percentage differences are relative to CAM-chem. Global mean differences are in legend. Values are for 2016.

Table 3.2: Global budgets of tropospheric ozone and OH.

Budget terms	GEOS-Chem	CAM-chem	Previous literature ^b
Tropospheric ozone burden (Tg)	350	342	340 (250-410)
O _x chemical production (Tg a ⁻¹)	5395	5052	4900 (3800-6900)
O _x chemical loss (Tg a ⁻¹)	4813	4465	4600 (3300-6600)
O _x deposition (Tg a ⁻¹)	878	967	
Ozone dry deposition (Tg a ⁻¹)	749	826	1000 (700-1500)
O _x STE (Tg a ⁻¹) ^c	341	380	500 (180-920)
O _x Lifetime (days)	23.0	23.7	22.3 (19.9-25.5)
Global OH (10 ⁶ molecule cm ⁻³) ^d	1.21	1.22	1.11 ± 0.16
N/S ratio	1.22	1.26	MMM: 1.28 ± 0.10; Obs.: 0.85 - 0.98
τ_{MCF} (a)	5.4	5.3	MMM: 5.7 ± 0.9; Obs.: 6.3 ± 0.4
Stratospheric ozone burden (Tg)	2744	2744	

^a. Annual mean values for 2016 from GEOS-Chem and CAM-chem in CESM2. The troposphere is defined by O₃ < 150 ppb (Young et al., 2013). The budget is for the odd oxygen (O_x) family to account for rapid cycling between O_x species: O_x ≡ O₃ + O + O(¹D) + NO₂ + 2NO₃ + HNO₃ + particulate nitrate + HNO₄ + 3N₂O₅ + organic nitrates + Criegee intermediates + XO + HOX + XNO₂ + 2XNO₃ + 2OIO + 2I₂O₂ + 3I₂O₃ + 4I₂O₄ + 2Cl₂O₂ + 2OClO, where X is Cl, Br, or I. CAM-chem does not include particulate nitrate or tropospheric halogen species.

^b. Means and ranges from Young et al. (2018) (33 models) for ozone and Naik et al. (2013) (16 ACCMIP models) for OH, for the year 2000. MMM: Multi-model mean. Obs.: Observation-derived estimates.

^c. Stratosphere-Troposphere Exchange (STE) is estimated from the residual of mass balance between tropospheric chemical production and loss, O_x deposition, and accumulation.

^d. Global annual mean air-mass-weighted OH concentration in the troposphere.

Table 3.3: Global tropospheric budget terms from different configurations of GEOS-Chem and CAM-chem in CESM2.

Simulation	GEOS-Chem				CAM-chem	
	Standard	No nitrate photolysis	No N ₂ O ₅ cloud uptake	No halogen chemistry ^b	Standard	Fast-JX photolysis
Ozone burden (Tg)	350	332	356	365	342	355
O _x chemical production (Tg a ⁻¹)	5395	4902	5473	5048	5052	5233
O _x chemical loss (Tg a ⁻¹)	4813	4425	4882	4542	4465	4469
O _x lifetime (days)	23.0	24.7	23.0	25.1	23.7	24.1
Global OH (10 ⁶ molecule cm ⁻³) ^c	1.21	1.06	1.20	1.32	1.22	1.22
NO _x burden (Gmol N)	8.66	8.25	8.97	9.61	9.03	8.36

^a. Refer to footnotes in Table 2. The Standard entries replicate those of Table 2.

^b. Zeroing out reaction rates for halogen reactions in the troposphere.

^c. Global annual mean air-mass-weighted OH concentration in the troposphere.

3.5 COMPARISONS TO OZONESONDES AND TO ATOM-1 AIRCRAFT CAMPAIGN

Figure 3.5 compares annual mean ozone vertical profiles simulated by GEOS-Chem and CAM-chem to ozonesonde observations for 2016 from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC), averaged across 9 regions following Tilmes et al. (2012). Both models match the observations well and are within 5–10 ppb of each other. GEOS-Chem has lower ozone at high northern latitudes because of halogen chemistry.

Figure 3.6 shows tropospheric profiles of OH, NO, and CO simulated by GEOS-Chem and CAM-chem over the oceans in comparison to observations from the ATom-1 campaign. Both models generally agree with OH observations within uncertainty. Both models fit NO observations within a factor of 2 in the northern hemisphere but have large underestimates in the southern hemisphere. This underestimate is a known model issue in previous offline GEOS-Chem simulations (Travis et al., 2020) and is not correctable by nitrate photolysis because particulate nitrate concentrations in the southern hemisphere are low (Shah et al., 2023). Observations show a NO increase in the upper troposphere of the southern hemisphere that is captured by CAM-chem but not GEOS-Chem. Previous work has shown that offline GEOS-Chem simulations capture this increase of NO in the upper troposphere (Shah et al., 2023). A sensitivity GEOS-Chem simulation without tropospheric halogen chemistry, as shown in Figure 6, also captures this increase. One possible reason is CESM2 not explicitly representing scavenging in convective updrafts for soluble gases such as HBr and HOBr. This would increase the formation of stable halogen nitrates in the upper troposphere where thermolysis and hydrolysis are slow (Wang et al., 2021).

Both models underestimate CO in the northern hemisphere which is a known issue attributed to excessive OH (Gaubert et al., 2020) or missing emissions of CO and its precursors (Park et al., 2021; Tang et al., 2023). CAM-chem has 10–20 ppb lower CO globally compared to GEOS-Chem that is likely driven by differences in OH. In the southern hemisphere the difference is driven by

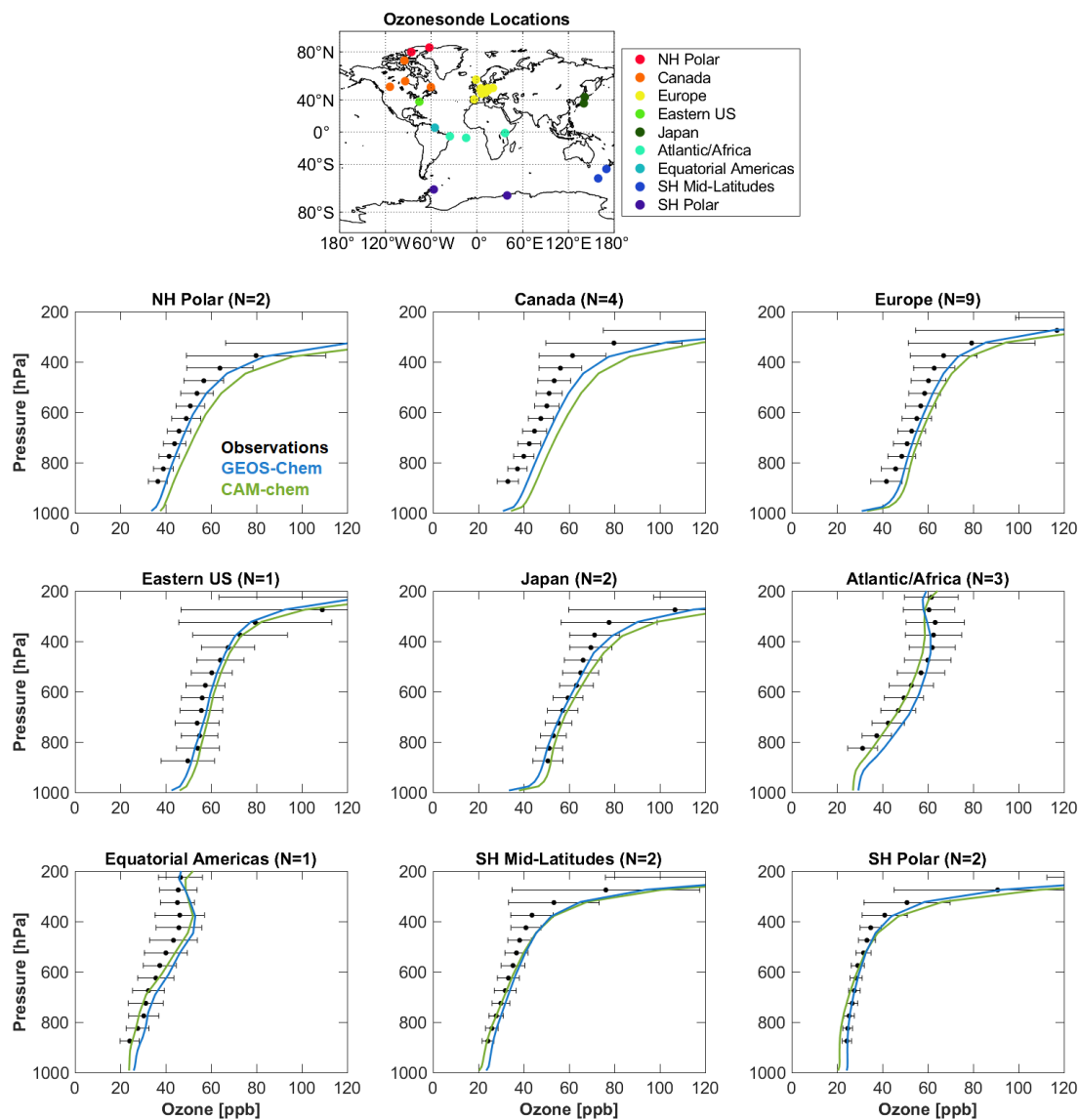


Figure 3.5: Comparison of GEOS-Chem and CAM-chem simulated annual mean ozone vertical profiles to 2016 ozonesonde observations. The regions average a number N of observing sites as given by Tilmes et al. (2012). Horizontal bars are standard deviations of the means across the N sites.

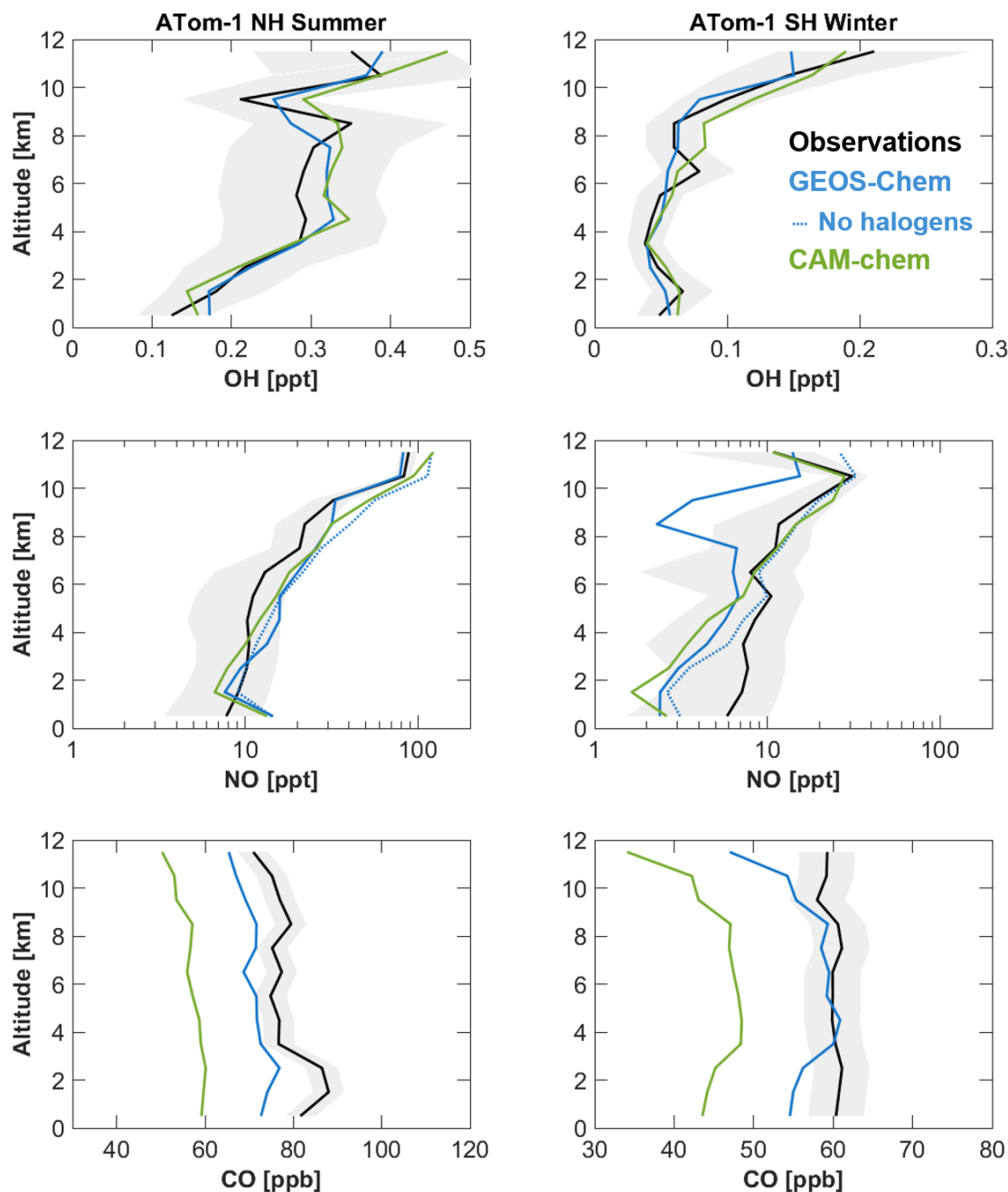


Figure 3.6: Median vertical profiles of OH, NO, and CO concentrations from the ATom-1 field campaign (July-August 2016) and from the GEOS-Chem and CAM-chem models within CESM2. Observations are separated between northern and southern hemispheres (NH and SH), filtered to remove influences from biomass burning ($\text{CH}_3\text{CN} > 200$ ppt; Travis et al., 2020) and binned in 1km intervals. Shaded areas correspond to the measurement accuracy.

improvements in isoprene oxidation in GEOS-Chem by Bates and Jacob (2019), which recycles OH through H-shift isomerization of isoprene-hydroxy-peroxy radicals under low-NO conditions, seen in observations by Wells et al. (2020). This leads to faster in situ isoprene oxidation and a higher CO yield. This is not included in CAM-chem's default MOZART-TS₁ mechanism used in this work but is included in the updated MOZART-TS₂ mechanism (Schwantes et al., 2020, 2021).

3.6 COMPARISON TO KORUS-AQ AIRCRAFT CAMPAIGN

We use comparison to observations from the KORUS-AQ campaign (May 1 to June 10, 2016) over the Seoul Metropolitan Area (SMA, 37–37.6°N, 126.6–127.7°E) as illustrative of a polluted atmosphere. Figure 3.7 shows median concentration profiles of oxidants and related species. Observations are compared to GEOS-Chem and CAM-chem sampled along the flight tracks, and to GEOS-Chem sensitivity simulations without particulate nitrate photolysis and without the nitrate correction applied in CESM2 for lack of scavenging in convective updrafts. Also shown in the Figure are vertical profiles from an offline regional GEOS-Chem simulation reported by Yang et al. (2023). Ozone vertical profiles in GEOS-Chem and CAM-chem are comparable and consistent with observations (Fig. 3.7a). Successful simulation of ozone in GEOS-Chem is contingent on particulate nitrate photolysis, offsetting the loss from halogen chemistry (Colombi et al., 2023; Yang et al., 2023). A previous ozone multi-model intercomparison with KORUS-AQ observations by Park et al. (2021) found GEOS-Chem without particulate nitrate photolysis to be too low, consistent with the results shown here. CAM-chem was the only model to successfully reproduce observed ozone in that intercomparison and this was attributed to its stratospheric ozone influx, but here GEOS-Chem uses the same dynamics and hence the same stratospheric influx. The success of CAM-chem in KORUS-AQ reflects instead its non-accounting of tropospheric halogen chemistry as a sink of ozone, which in GEOS-Chem needs to be compensated by particulate nitrate photolysis.

Particulate nitrate photolysis increases free tropospheric NO_x and ozone production but this depends on the nitrate concentration. CAM-chem does not simulate nitrate. Because GEOS-Chem nitrate is not removed in convective updrafts in the CESM2 environment, our standard implementation within CESM2 corrects nitrate using the same photolytic sink that CAM-chem applies for SOA with a rate of $0.0004 \times J_{\text{NO}_2}$ and no products (Hodzic et al., 2015, 2016) to avoid buildup in the upper troposphere. Nitrate in GEOS-Chem within CESM2 without this correction (dotted line

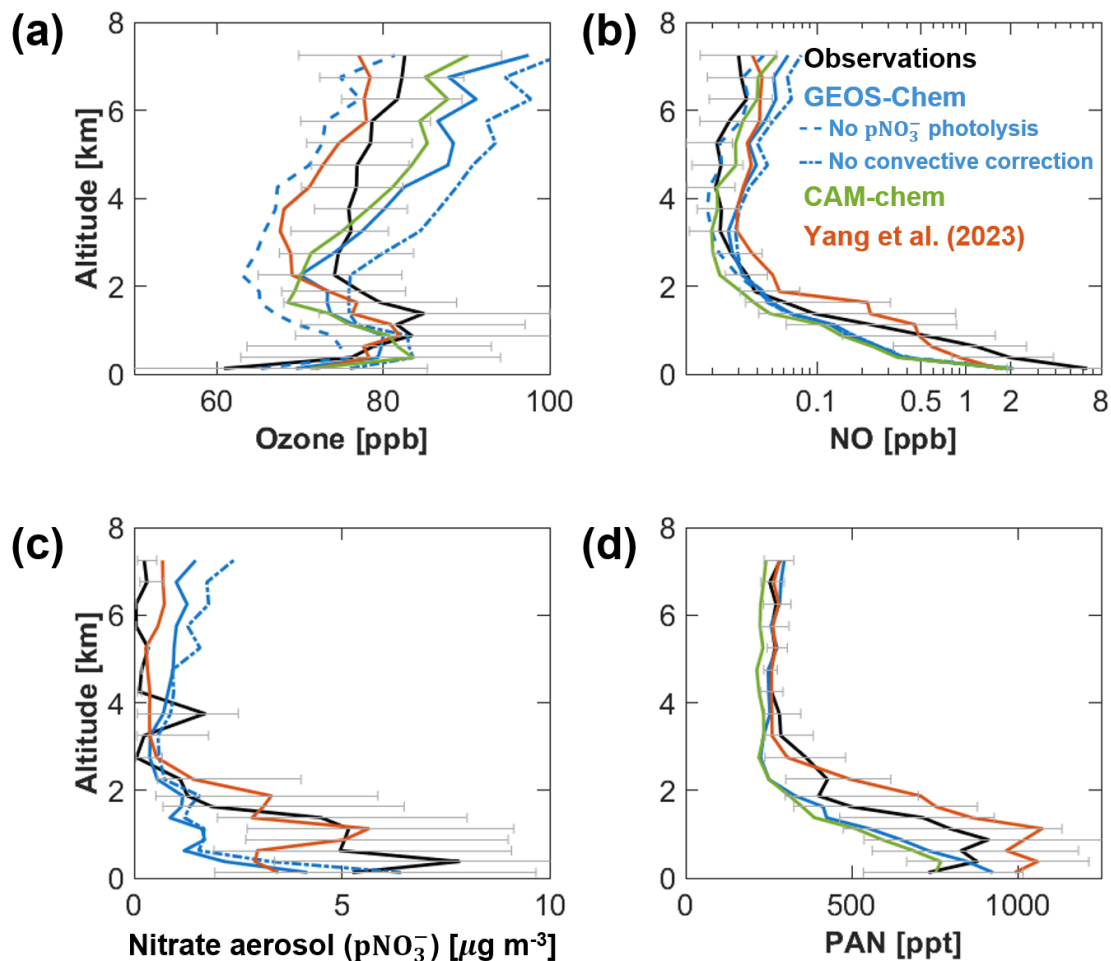


Figure 3.7: Median tropospheric vertical profiles of species concentrations over the Seoul Metropolitan Area (SMA; 37–37.6°N, 126.6–127.7°E) during the KORUS-AQ aircraft campaign in May-June 2016. Observations are compared to GEOS-Chem and CAM-chem simulations within CESM2, and to the offline GEOS-Chem simulation reported by Yang et al. (2023). Results from a GEOS-Chem sensitivity simulation with no particulate nitrate photolysis and no correction for scavenging of nitrate in wet convective updrafts are also shown. The vertical profiles are constructed by binning data into 0.25 km intervals below 2 km altitude and 0.5 km intervals above 2 km altitude. Horizontal bars represent the interquartile range of the observations in the given vertical bin.

in Fig. 3.7c) is overestimated above 3km which leads to excessive effects of particulate nitrate photolysis including overestimation of ozone (Fig. 3.7a) and NO (Fig. 7b). The offline GEOS-Chem simulation by Yang et al. (2023) explicitly scavenges nitrate in convective updrafts and does not overestimate particulate nitrate or NO. A solution would be to replace CESM convective transport with the GEOS-Chem offline convective transport and scavenging module using archived CESM convective mass fluxes, and this has been done before when coupling GEOS-Chem to the GEOS and BCC ESMs which had the same problem of not scavenging water-soluble species in convective updrafts (Yu et al., 2018; Lu et al., 2020). A more comprehensive solution would be to include scavenging of water-soluble species in the CESM2 convection scheme. This is implemented for MAM aerosols (Wang et al., 2013) but not for gas-phase species or aerosols only represented in GEOS-Chem, including nitrate.

The simulations of particulate nitrate and peroxyacetylnitrate (PAN) within CESM show a sharp drop of concentrations with altitude above the surface, whereas the observations and the offline GEOS-Chem simulation of Yang et al. (2023) show a mixed layer structure extending to 1–2 km altitude. This likely reflects a bias in the CESM2 boundary layer mixing scheme that would need to be investigated further. Boundary layer mixing in the offline GEOS-Chem model is a standard non-local scheme from J. Lin and McElroy (2014). The PAN simulations in GEOS-Chem and CAM-chem otherwise agree closely, indicating similar production from VOC chemistry, and are lower than the offline GEOS-Chem simulation which includes additional emissions of volatile chemical products (VCPs) as a source of acetaldehyde leading to PAN production (Yang et al., 2023).

3.7 CONCLUSIONS

GEOS-Chem has been implemented as an atmospheric chemistry module in the NCAR Community Earth System Model (CESM2) to serve as alternative to CAM-chem and contribute to the MU-

SICA vision of plug-and-play modularization of atmospheric chemistry within CESM (Pfister et al., 2020). Here we presented an intercomparison and evaluation with observations of tropospheric oxidant simulations with these two modules. The intercomparison covered the full year of 2016, allowing evaluation with the ATom-1 aircraft campaign over the remote Pacific and Atlantic, and the KORUS-AQ aircraft campaign over the Seoul Metropolitan Area (SMA). Both GEOS-Chem and CAM-chem used the same emissions processed through HEMCO (Lin et al., 2021) and the same coupling to other CESM2 modules.

GEOS-Chem uses the Fast-JX scheme for photolysis while CAM-chem uses a lookup table based on TUV. Both schemes agree to within 10% when compared to J_{NO_2} and $J_{\text{O}_1\text{D}}$ photolysis frequencies observations in ATom-1, but observations in KORUS-AQ show that CAM-chem overestimates $J_{\text{O}_1\text{D}}$ while GEOS-Chem does not. One major difference is that TUV does not account for extinction by aerosols while Fast-JX does. We implemented Fast-JX in CAM-chem and find that it resolves most of the photolysis differences with GEOS-Chem.

Global tropospheric ozone budget terms in GEOS-Chem and CAM-chem agree within 10%, compared to a much wider spread in the literature and due in part to canceling effects. Differences between the two models are mostly driven by aerosol nitrate photolysis, N_2O_5 uptake in clouds, and tropospheric halogen chemistry, all of which are included in GEOS-Chem but not in CAM-chem. Aerosol nitrate photolysis in GEOS-Chem produces NO_x and enhances ozone production, compensating for losses from N_2O_5 uptake in clouds and tropospheric halogen chemistry. Annual mean ozone concentrations agree within 5 ppb between GEOS-Chem and CAM-chem almost everywhere. Lower ozone deposition to the oceans in GEOS-Chem results in higher surface ozone at southern mid-latitudes. Tropospheric halogen chemistry results in lower ozone at high northern latitudes. Tropospheric NO_x in GEOS-Chem is higher than CAM-chem in the tropics due to nitrate photolysis, and lower at high latitudes due to N_2O_5 uptake by cloud and formation of halogen nitrates. The global mean tropospheric OH concentration is identical between the two models but

there are large differences over the continents driven by photolysis and by isoprene chemistry.

Both GEOS-Chem and CAM-chem show good agreement with annual mean ozonesonde observations over the range of latitudes. Comparison to ATom-1 observations in July-August 2016 shows good agreement for OH concentrations in both the northern and southern hemispheres (NH and SH) within the measurement accuracy, and for NO_x in the NH, but NO_x in the SH is underestimated. GEOS-Chem shows a depletion of NO_x in the SH upper troposphere that is due to formation of halogen nitrates and is not seen in the observations. However, the offline GEOS-Chem simulation does not show this problem. One issue in CESM2 is the lack of scavenging of water-soluble species including halogen radical reservoirs in convective updrafts. Both GEOS-Chem and CAM-chem underestimate CO in the NH but CAM-chem is consistently lower than GEOS-Chem due to higher OH in the NH and suppression of CO production from isoprene oxidation in the SH.

Comparison with KORUS-AQ aircraft observations allowed model evaluation for polluted conditions. Vertical ozone profiles in GEOS-Chem and CAM-chem agree well with observations, which in GEOS-Chem is contingent on the NO_x source from particulate nitrate photolysis. Lack of scavenging of GEOS-Chem aerosols such as nitrate and its gas phase precursors in convective updrafts is a major shortcoming in CESM2 that hinders proper representation of nitrate and other aerosols in the upper troposphere. Comparison of peroxyacetylnitrate (PAN) shows good agreement between GEOS-Chem and CAM-chem and with observations, indicating consistency in the VOC chemistry producing PAN. However, the decrease of PAN and particulate nitrate mixing ratios with altitude are much sharper than observed and simulated by the offline GEOS-Chem model, implying insufficient boundary layer mixing in CESM2.

Overall, we have shown that GEOS-Chem provides a high-quality simulation of tropospheric oxidant chemistry in CESM2 and can contribute modules for alternative representations of atmospheric chemistry to serve the MUSICA vision.

3.8 CODE AVAILABILITY

A fork of an alpha version (cam6_3_095) of the Community Atmosphere Model (CAM) including GEOS-Chem is available at <https://github.com/CESM-GC/CAM> and is used in this work. CAM-chem using HEMCO for emissions is implemented in the mainline CAM code as of cam6_3_118 (https://github.com/ESCOMP/CAM/tree/cam6_3_118). GEOS-Chem within CESM2 is implemented in the mainline CAM code as of cam6_3_147 (https://github.com/escomp/cam/tree/cam6_3_147).

3.9 AUTHOR CONTRIBUTIONS AND ACKNOWLEDGEMENTS

LKE, SDE, and DJJ conceived of the project, acquired funding, and supervised the work. HL, EWL, XF, SDE, and TMF performed software development. HL, LKE, LHY, XF, RD, SZ, YT, MK, NKC, and DJJ analyzed the model results. SZ prepared the KORUSv5 emission inventory for input into HEMCO and CESM2. HL performed the visualization and preparation of the original draft. All authors contributed to review and editing of the manuscript.

This work was supported by the Atmospheric Chemistry Program of the US National Science Foundation. We would like to acknowledge high-performance computing support from Cheyenne: HPE/SGI ICE XA System (<https://doi.org/10.5065/D6RX99HX>; Hart, 2021) and Derecho: HPE Cray EX System (<https://doi.org/10.5065/qx9a-pg09>), provided by NCAR's Computational and Information Systems Laboratory (CISL) and sponsored by the National Science Foundation.



Appendix A: Supporting information for
Chapter 2

Table A.1: List of species within each category for mass balance calculations.

Odd oxygen (Ox)	O ₃ , NO ₂ , NO ₃ , PAN, PPN, MPAN, HNO ₄ , N ₂ O ₅ , HNO ₃ , BrO, HOBr, BrNO ₂ , BrNO ₃ , MPN, ETHLN, MVKN, MCRHN, MCRHNB, PROPNN, R ₄ N ₂ , PRN ₁ , PRPN, R ₄ N ₁ , HONIT, MONITS, MONITU, OLND, OLNN, IHN ₁ , IHN ₂ , IHN ₃ , IHN ₄ , INPB, INPD, ICN, IDN, ITCN, ITHN, ISOPNOO ₁ , ISOPNOO ₂ , INO ₂ B, INO ₂ D, INA, IDHNBOO, IDHNDOO ₁ , IDHNDOO ₂ , IHPNBOO, IHPNDOO, ICNOO, IDNOO, MACRNO ₂ , ClO, HOCl, ClNO ₂ , ClNO ₃ , Cl ₂ O ₂ , OCIO, O, O ₁ D, IO, HOY, IONO, IONO ₂ , OIO, I ₂ O ₂ , I ₂ O ₃ , I ₂ O ₄
Reactive Nitrogen	HNO ₂ , HNO ₃ , HNO ₄ , MPAN, NIT, NO, NO ₂ , NO ₃ , N ₂ O ₅ , MPN, PAN, PPN, N ₂ O, MENO ₃ , ETNO ₃ , IPRNO ₃ , NPRNO ₃ , AONITA, ETHN, BZPAN, NPHEN
Chlorine	CH ₂ ICl, CH ₃ CCl ₃ , CCl ₄ , CH ₂ Cl ₂ , Cl ₂ O ₂ , ICl, CH ₃ Cl, ClOO, OCIO, BrCl, Cl ₂ , ClNO ₂ , ClNO ₃ , HOCl, HCl, Cl, ClO
Bromine	CH ₂ IBr, BrNO ₂ , CH ₂ Br ₂ , IBr, CHBr ₃ , CH ₃ Br, Br ₂ , BrCl, BrNO ₃ , HOBr, BrO, BrSALA, BrSALC, Br, HBr

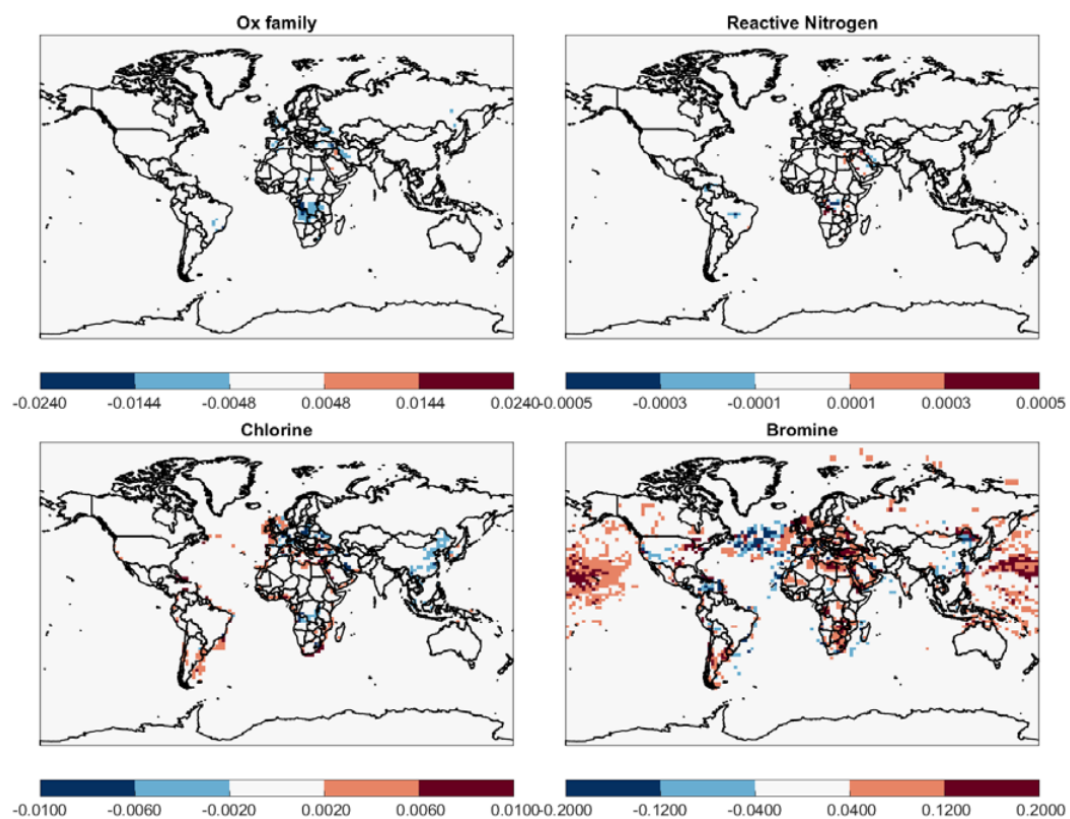


Figure A.1: Mass imbalance in percent (%) of total odd oxygen (O_x), reactive nitrogen, chlorine, and bromine atoms (as defined in Table A.1) for the adaptive solver in its standard configuration compared with the full mechanism solver, at a snapshot in time on July 1, 2014, 00:40 UTC, at surface.

Table A.2: Computational performance and accuracy of the adaptive solver for different configurations. The solver wall clock times and all species median relative root mean squared (RRMS) error percentages are obtained from a 14-day simulation of the GEOS-Chem model with KPP 3.0.0.

Configuration	Solver wall clock time [s] (Speed-up)	All Species median RRMS Error [%]		
		PBL	Free Trop.	Stratosphere
Dynamics timestep: 600s, Chemistry timestep: 1200s				
Full solver	3666.8			
Adaptive solver $\alpha_{OH} = 5 \times 10^{-5}$ $\alpha_{NO2} = 1 \times 10^{-4}$	2487.0 (32%)	0.80	0.92	1.28
Adaptive solver $\alpha_{OH} = 5 \times 10^{-5}$ $\alpha_{NO3} = 1 \times 10^{-4}$	2559.1 (30%)	0.56	0.80	1.19
Adaptive solver $\alpha_{OH} = 5 \times 10^{-5}$ $\alpha_{NO2} = 1 \times 10^{-4}$ Keep halogens active	2824.40 (23%)	0.59	0.77	0.83
Adaptive solver $\alpha_{OH} = 5 \times 10^{-5}$ $\alpha_{NO2} = 1 \times 10^{-4}$ In troposphere only Full solver in stratosphere	2922.21 (20%)	0.81	0.92	0.37
Dynamics timestep: 300s, Chemistry timestep: 600s				
Full solver	4684.71			
Adaptive solver $\alpha_{OH} = 5 \times 10^{-5}$ $\alpha_{NO2} = 1 \times 10^{-4}$	3605.25 (23%)	0.58	0.54	0.61
Adaptive solver $\alpha_{OH} = 1 \times 10^{-4}$ $\alpha_{NO2} = 2 \times 10^{-4}$	3369.28 (28%)	0.95	0.86	0.86
Dynamics timestep: 900s, Chemistry timestep: 1800s				
Full solver	3175.85			
Adaptive solver $\alpha_{OH} = 5 \times 10^{-5}$ $\alpha_{NO2} = 1 \times 10^{-4}$	2067.26 (35%)	1.23	1.48	2.31
Adaptive solver $\alpha_{OH} = 2.5 \times 10^{-5}$ $\alpha_{NO2} = 5 \times 10^{-5}$	2146.56 (32%)	0.67	0.91	1.61

Table A.3: List of species within each category for accuracy calculations based on the GEOS-Chem benchmarking process (http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_benchmarking).

Oxidants	O ₃ , CO, OH, NO, NO ₂ , NO ₃
RO_y	H ₂ O ₂ , H, H ₂ , H ₂ O, HO ₂ , O ₁ D, OH
Nitrogen	HNO ₂ , HNO ₄ , MPAN, HNO ₃ T, NO, NO ₂ , NO ₃ , N ₂ O ₅ , MPN, PAN, PPN, N ₂ O, NH ₃ , NH ₄ , MENO ₃ , ETNO ₃ , IPRNO ₃ , NPRNO ₃
Primary Organics	EOH, MOH, ISOP, MTPA, MTPO, LIMO, ALK ₄ , BENZ, CH ₄ , C ₂ H ₆ , C ₃ H ₈ , PRPE, TOLU, XYLE
Secondary Organics	ACTA, ALD ₂ , CH ₂ O, MACR, HPALD ₁ , HPALD ₂ , HPALD ₃ , HPALD ₄ , IEPOXA, IEPOXB, IEPOXD, ACET, HAC, MEK, MVK, IHN ₁ , IHN ₂ , IHN ₃ , IHN ₄ , GLYC, GLYX, HCOOH, MAP, RCHO, MP
Sulfur	DMS, MSA, OCS, SO ₂ , SO ₄
Bromine	Br, Br ₂ , BrCl, BrNO ₂ , BrNO ₃ , BrO, CH ₃ Br, CH ₂ Br ₂ , CHBr ₃ , HOBr, HBr
Chlorine	Cl, ClO, Cl ₂ , Cl ₂ NO ₂ , ClOO, ClNO ₂ , ClNO ₃ , CCl ₄ , CH ₃ Cl, CH ₂ Cl ₂ , CH ₃ CCl ₃ , CHCl ₃ , HOCl, HCl, OClO
Iodine	I, I ₂ , IBr, ICl, IO, IONO, IONO ₂ , CH ₃ I, CH ₂ I ₂ , CH ₂ ICl, CH ₂ IBr, HI, HOI, OIO

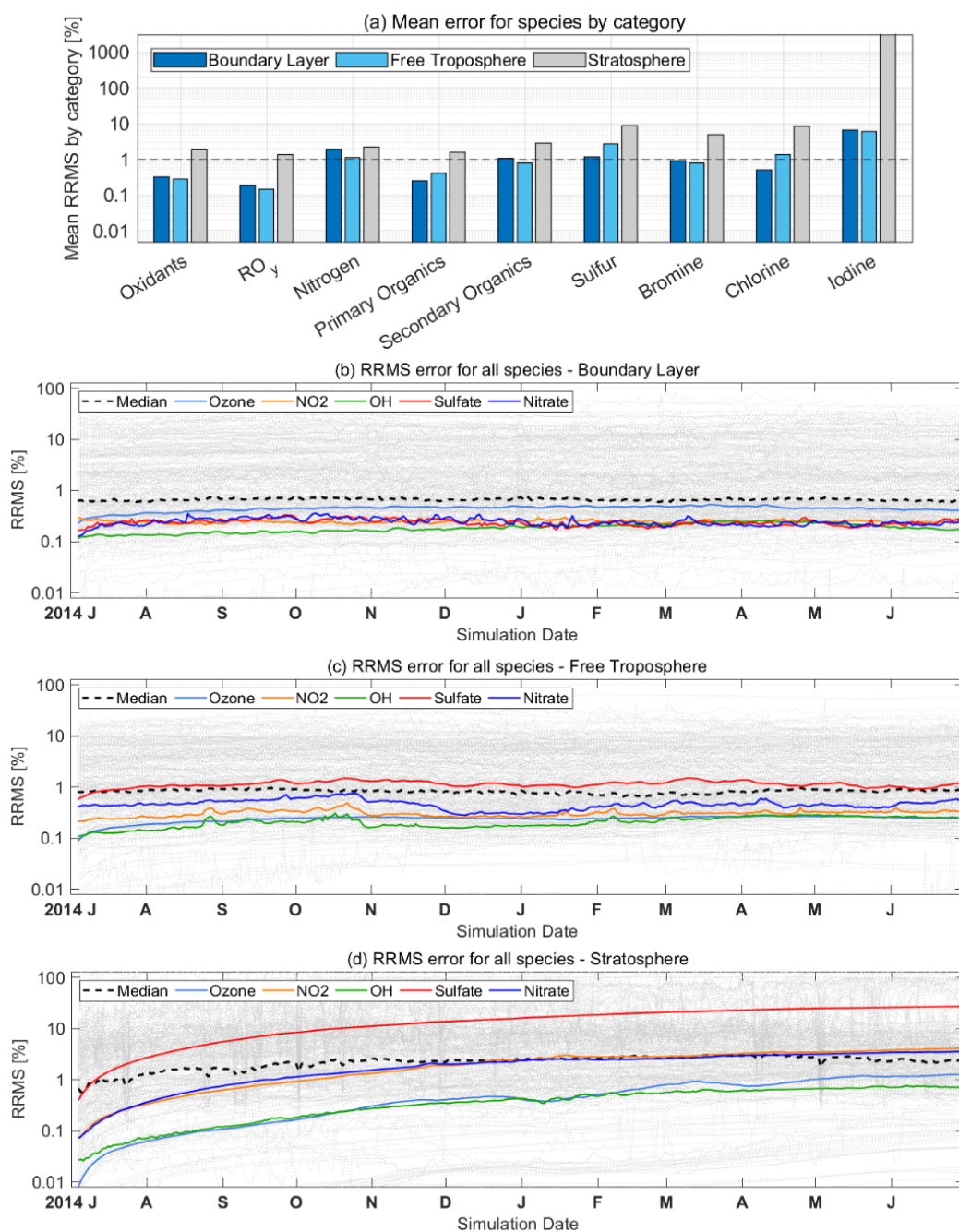


Figure A.2: Accuracy of the adaptive solver in a 1-year GEOS-Chem simulation, keeping halogen species active. Panel (a) shows the mean RRMS errors for species in different categories at the end of the 1-year simulation starting on July 1, 2014. The categories are as defined in the standard GEOS-Chem benchmarking output diagnostics (Table A.2). Panels (b)-(d) show the time evolution of RRMS errors for all species in the boundary layer, free troposphere, and stratosphere, respectively, with colored lines for species of particular interest.

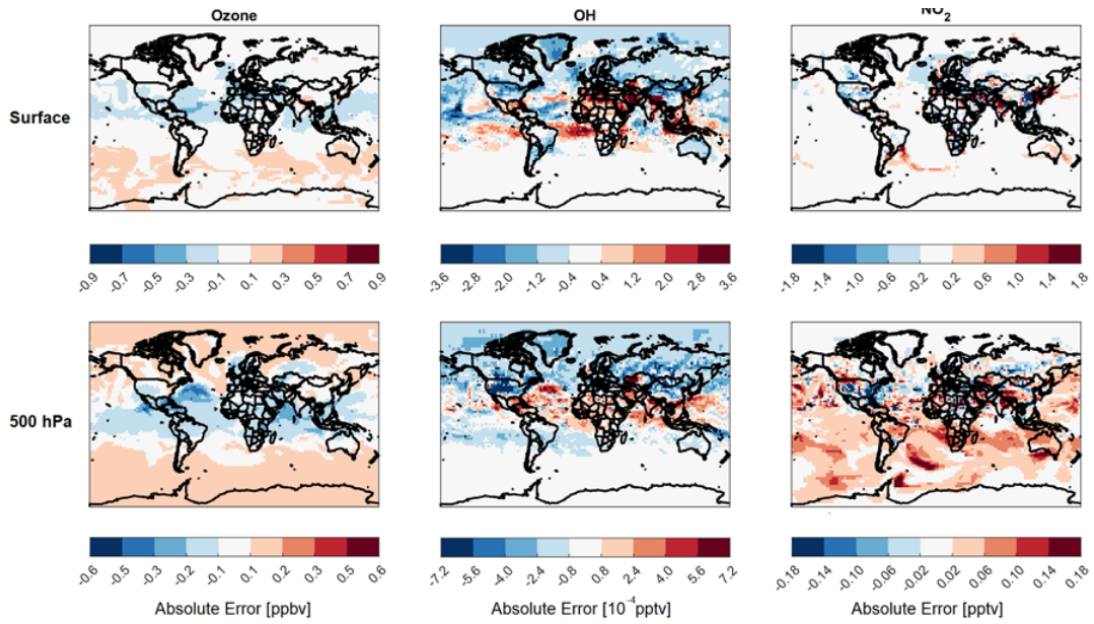


Figure A.3: Accuracy of the adaptive solver for ozone, OH, and NO_2 global distributions (Absolute error). This figure shows the equivalent of Figure 2.5, but plotting absolute errors of the adaptive solver relative to the full solver for the daily mean concentrations on July 1, 2015, the last day of the 1-year simulation.

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.
- Alexander, B., Park, R. J., Jacob, D. J., Li, Q. B., Yantosca, R. M., Savarino, J., Lee, C. C. W., and Thiemens, M. H.: Sulfate formation in sea-salt aerosols: Constraints from oxygen isotopes, *J. Geophys. Res. Atmos.*, 110(D10), D10307, <https://doi.org/10.1029/2004JD005659>, 2005.
- Alvanos, M., and Christoudias, T.: GPU-accelerated atmospheric chemical kinetics in the ECHAM/MESSy (EMAC) Earth system model (version 2.52), *Geosci. Model Dev.*, 10, 3679–3693, <https://doi.org/10.5194/gmd-10-3679-2017>, 2017.
- Amos, H. M., Jacob, D. J., Holmes, C. D., Fisher, J. A., Wang, Q., Yantosca, R. M., Corbitt, E. S., Galarneau, E., Rutter, A. P., Gustin, M. S., Steffen, A., Schauer, J. J., Graydon, J. A., Louis, V. L. St., Talbot, R. W., Edgerton, E. S., Zhang, Y., and Sunderland, E. M.: Gas-particle partitioning of atmospheric Hg(II) and its effect on global mercury deposition, *Atmos. Chem. Phys.*, 12, 591–603, <https://doi.org/10.5194/acp-12-591-2012>, 2012.
- 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, 955–966, <https://doi.org/10.1029/2000GB001382>, 2001.
- Balkanski, Y. J., Jacob, D. J., Gardner, G. M., Graustein, W. C., and Turekian, K. K.: Transport and residence times of tropospheric aerosols inferred from a global three-dimensional simulation of Pb-210, *J. Geophys. Res.*, 98(D11), 20573–20586, <https://doi.org/10.1029/93JD02456>, 1993.
- Bates, K. H., and Jacob, D. J.: A new model mechanism for atmospheric oxidation of isoprene: Global effects on oxidants, nitrogen oxides, organic products, and secondary organic aerosol, *Atmos. Chem. Phys.*, 19(14), 9613–9640, <https://doi.org/10.5194/acp-19-9613-2019>, 2019.
- Bates, K. H., and Jacob, D. J.: An expanded definition of the odd oxygen family for tropospheric ozone budgets: Implications for ozone lifetime and stratospheric influence, *Geophys. Res. Lett.*, 47(4), e2019GL084486, <https://doi.org/10.1029/2019GL084486>, 2020.
- Bates, K. H., Jacob, D. J., Li, K., Ivatt, P. D., Evans, M. J., Yan, Y., and Lin, J.: Development and evaluation of a new compact mechanism for aromatic oxidation in atmospheric models, *Atmos. Chem. Phys.*, 21(24), 18351–18374, <https://doi.org/10.5194/acp-21-18351-2021>, 2021.
- 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, 23073–23095, <https://doi.org/10.1029/2001JD000807>, 2001.

Bian, H., and Prather, M. J.: Fast-J2: Accurate simulation of stratospheric photolysis in global chemical models. *Journal of Atmospheric Chemistry*, *J. Atmos. Chem.*, 41(3), 281–296, <https://doi.org/10.1023/A:1014980619462>, 2002.

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, 561–587, <https://doi.org/10.1029/97GB02266>, 1997.

Brasseur, G. P., and Jacob, D. J.: Modeling of atmospheric chemistry, Cambridge University Press, <https://doi.org/10.1017/9781316544754>, 2017.

Brown, P. N., Byrne, G. D., and Hindmarsh, A. C.: VODE: A variable-coefficient ODE solver, *SIAM Journal on Scientific and Statistical Computing*, 10(5), 1038–1051, <https://doi.org/10.1137/0910062>, 1989.

Brown-Steiner, B., Selin, N. E., Prinn, R., Tilmes, S., Emmons, L., Lamarque, J.-F., and Cameron-Smith, P.: Evaluating simplified chemical mechanisms within present-day simulations of the community Earth system model version 1.2 with CAM4 (CESM1.2 CAM-chem): MOZART-4 vs. reduced hydrocarbon vs. super-fast chemistry, *Geosci. Model Dev.*, 11(10), 4155–4174, <https://doi.org/10.5194/gmd-11-4155-2018>, 2018.

Burkholder, J. B., Sander, S. P., Abbatt, J., Barker, J. R., Cappa, C., Crounse, J. D., Dibble, T. S., Huie, R. E., Kolb, C. E., Kurylo, M. J., Orkin, V. L., Percival, C. J., Wilmouth, D. M., and Wine, P. H.: Chemical kinetics and photochemical data for use in atmospheric studies, Evaluation No. 19, Tech. Rep. JPL Publication 19-5, Jet Propulsion Laboratory, Pasadena, CA, 2020.

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, 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, 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, *Mon. Weather Rev.*, 141, 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 GO-

CART Model and Comparisons with Satellite and Sun Photometer Measurements, *J. Atmos. Sci.*, 59, 461–483, [https://doi.org/10.1175/1520-0469\(2002\)059<0461:TAOTFT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2), 2002.

Colombi, N. K., Jacob, D. J., Yang, L. H., Zhai, S., Shah, V., Grange, S. K., Yantosca, R. M., Kim, S., and Liao, H.: Why is ozone in South Korea and the Seoul metropolitan area so high and increasing?, *Atmos. Chem. Phys.*, 23, 4031–4044, <https://doi.org/10.5194/acp-23-4031-2023>, 2023.

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

Crawford, J. H., Ahn, J.-Y., Al-Saadi, J., Chang, L., Emmons, L. K., Kim, J., Lee, G., Park, J.-H., Park, R. J., Woo, J. H., Song, C.-K., Hong, J.-H., Hong, Y.-D., Lefer, B. L., Lee, M., Lee, T., Kim, S., Min, K.-E., Yum, S. S., Shin, H. J., Kim, Y.-W., Choi, J.-S., Park, J.-S., Szykman, J. J., Long, R. W., Jordan, C. E., Simpson, I. J., Fried, A., Dibb, J. E., Cho, S., and Kim, Y. P.: The Korea–United States Air Quality (KORUS-AQ) field study, *Elem. Sci. Anth.*, 9, 00163, <https://doi.org/10.1525/elementa.2020.00163>, 2021.

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.

Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G. R.: The kinetic preprocessor KPP—a software environment for solving chemical kinetics, *Computers & Chemical Engineering*, 26(11), 1567–1579, [https://doi.org/10.1016/S0098-1354\(02\)00128-X](https://doi.org/10.1016/S0098-1354(02)00128-X), 2002.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), *J. Adv. Model. Earth Sy.*, 12(2), e2019MS001916, <https://doi.org/10.1029/2019MS001916>, 2020.

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, available at: <https://ntrs.nasa.gov/api/citations/20180005253/downloads/20180005253.pdf> (last access: 2 September 2021), 2015.

Dawson, M. L., Guzman, C., Curtis, J. H., Acosta, M., Zhu, S., Dabdub, D., Conley, A., West, M., Riemer, N., and Jorba, O.: Chemistry Across Multiple Phases (CAMP) version 1.0: an inte-

grated multiphase chemistry model, *Geosci. Model Dev.*, 15, 3663–3689, <https://doi.org/10.5194/gmd-15-3663-2022>, 2022.

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.

Djouad, R., and Sportisse, B.: Partitioning techniques and lumping computation for reducing chemical kinetics. APLA: An automatic partitioning and lumping algorithm, *Appl. Numer. Math.*, 43(4), 383–398, [https://doi.org/10.1016/S0168-9274\(02\)00111-3](https://doi.org/10.1016/S0168-9274(02)00111-3), 2002.

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, <https://doi.org/10.1016/j.atmosenv.2014.02.001>, 2014.

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.

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., Chen, Q., 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.*, 14, 3741–3768, <https://doi.org/10.5194/gmd-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 atmospheric chemistry models, *Geosci. Model Dev.*, 4, 419–433, <https://doi.org/10.5194/gmd-4-419-2011>, 2011.
- Fritz, T. M., Eastham, S. D., Emmons, L. K., Lin, H., Lundgren, E. W., Goldhaber, S., Barrett, S. R. H., and Jacob, D. J.: Implementation and evaluation of the GEOS-Chem chemistry module version 13.1.2 within the Community Earth System Model v2.1, *Geosci. Model Dev.*, 15, 8669–8704, <https://doi.org/10.5194/gmd-15-8669-2022>, 2022.
- Gaubert, B., Emmons, L. K., Raeder, K., Tilmes, S., Miyazaki, K., Arellano Jr., A. F., Elguindi, N., Granier, C., Tang, W., Barré, J., Worden, H. M., Buchholz, R. R., Edwards, D. P., Franke, P., Anderson, J. L., Saunio, M., Schroeder, J., Woo, J.-H., Simpson, I. J., Blake, D. R., Meinardi, S., Wennberg, P. O., Crounse, J., Teng, A., Kim, M., Dickerson, R. R., He, H., Ren, X., Pusede, S. E., and Diskin, G. S.: Correcting model biases of CO in East Asia: impact on oxidant distributions during KORUS-AQ, *Atmos. Chem. Phys.*, 20, 14617–14647, <https://doi.org/10.5194/acp-20-14617-2020>, 2020.
- Ge, C., Wang, J., Carn, S., Yang, K., Ginoux, P., and Krotkov, N.: Satellite-based global volcanic SO₂ emissions and sulfate direct radiative forcing during 2005–2012, *J. Geophys. Res.-Atmos.*, 121, 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, 317–328, <https://doi.org/10.1002/jgrg.20042>, 2013.
- 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, 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, 1097, <https://doi.org/10.1029/2003GB002079>, 2003.
- Gong, W., and Cho, H.-R.: A numerical scheme for the integration of the gas-phase chemical rate equations in three-dimensional atmospheric models, *Atmos. Environ. A: General Topics*, 27(14), 2147–2160, [https://doi.org/10.1016/0960-1686\(93\)90044-Y](https://doi.org/10.1016/0960-1686(93)90044-Y), 1993.
- 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, available at: <http://www.aero.jussieu.fr/projet/ACCENT/POET.php> (last access: 23 July 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.
- 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.

- Hairer, E., and Wanner, G.: Solving ordinary differential equations II. Stiff and differential-algebraic problems, Springer, 1991.
- Hakami, A., Henze, D. K., Seinfeld, J. H., Singh, K., Sandu, A., Kim, S., Byun, and Li, Q.: The Adjoint of CMAQ, *Environ. Sci. Technol.*, 41(22), 7807–7817, <https://doi.org/10.1021/es070944p>, 2007.
- Hall, S. R., Ullmann, K., Prather, M. J., Flynn, C. M., Murray, L. T., Fiore, A. M., Correa, G., Strode, S. A., Steenrod, S. D., Lamarque, J.-F., Guth, J., Josse, B., Flemming, J., Huijnen, V., Abraham, N. L., and Archibald, A. T.: Cloud impacts on photochemistry: building a climatology of photolysis rates from the Atmospheric Tomography mission, *Atmos. Chem. Phys.*, 18, 16809–16828, <https://doi.org/10.5194/acp-18-16809-2018>, 2018.
- Harris, L. M. and Lin, S.-J.: A Two-Way Nested Global-Regional Dynamical Core on the Cubed-Sphere Grid, *Mon. Weather Rev.*, 141, 283–306, <https://doi.org/10.1175/MWR-D-11-00201.1>, 2013.
- Harris, L. M., Lin, S.-J., and Tu, C.: High-Resolution Climate Simulations Using GFDL Hi-RAM with a Stretched Global Grid, *J. Climate*, 29, 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 (GFASvo) by MACC, ECMWF Technical Memoranda, 628, 1–47, <https://doi.org/10.21957/2m000mza9>, 2010.
- Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, *Atmos. Chem. Phys.*, 7, 2413–2433, <https://doi.org/10.5194/acp-7-2413-2007>, 2007.
- 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, <https://doi.org/10.1109/MCISE.2004.1255817>, 2004.
- Hodzic, A., Madronich, S., Kasibhatla, P. S., Tyndall, G., Aumont, B., Jimenez, J. L., Lee-Taylor, J., and Orlando, J.: Organic photolysis reactions in tropospheric aerosols: effect on secondary organic aerosol formation and lifetime, *Atmos. Chem. Phys.*, 15, 9253–9269, <https://doi.org/10.5194/acp-15-9253-2015>, 2015.
- Hodzic, A., Kasibhatla, P. S., Jo, D. S., Cappa, C. D., Jimenez, J. L., Madronich, S., and Park, R. J.: Rethinking the global secondary organic aerosol (SOA) budget: stronger production, faster removal, shorter lifetime, *Atmos. Chem. Phys.*, 16, 7917–7941, <https://doi.org/10.5194/acp-16-7917-2016>, 2016.
- 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 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.

- Holmes, C. D., Bertram, T. H., Confer, K. L., Graham, K. A., Ronan, A. C., Wirks, C. K., and Shah, V.: The role of clouds in the tropospheric NO_x cycle: A new modeling approach for cloud chemistry and its global implications, *Geophys. Res. Lett.*, 46(9), 4980–4990, <https://doi.org/10.1029/2019GL081990>, 2019.
- Hu, L., Jacob, D. J., Liu, X., Zhang, Y., Kim, P. S., Sulprizio, M. P., and Yantosca, R. M.: Global budget of tropospheric ozone: Evaluating recent model advances with satellite (OMI), aircraft (IAGOS), and ozonesonde observations, *Atmos. Environ.*, 167, 323–334, <https://doi.org/10.1016/j.atmosenv.2017.08.036>, 2017.
- Hu, L., 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.
- 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, <https://doi.org/10.5194/acp-12-7779-2012>, 2012.
- Jacobson, M. Z., Computation of global photochemistry with SMVGEAR II, *Atmos. Environ.*, 29(18), 2541–2546, [https://doi.org/10.1016/1352-2310\(95\)00194-4](https://doi.org/10.1016/1352-2310(95)00194-4), 1995.
- Jacobson, M. Z., and Turco, R. P.: SMVGEAR: A sparse-matrix, vectorized gear code for atmospheric models, *Atmos. Environ.*, 28(2), 273–284, [https://doi.org/10.1016/1352-2310\(94\)90102-3](https://doi.org/10.1016/1352-2310(94)90102-3), 1994.
- 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 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, <https://doi.org/10.2788/14102>, 2012.
- Jöckel, P.: Technical note: Recursive rediscritisation 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.
- Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), *Geosci. Model Dev.*, 3, 717–752, <https://doi.org/10.5194/gmd-3-717-2010>, 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, <https://doi.org/10.5194/os-6-913-2010>, 2010.

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. and Evans, M. J.: Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v1.0, *Geosci. Model Dev.*, 12, 1209–1225, <https://doi.org/10.5194/gmd-12-1209-2019>, 2019.

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.

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, *J. Adv. Model. Earth Sy.*, 13(4), e2020MS002413, <https://doi.org/10.1029/2020MS002413>, 2021.

Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D., and Tessum, C. W.: Toward stable, general machine-learned models of the atmospheric chemical system, *J. Geophys. Res. Atmos.*, 125(23), e2020JD032759, <https://doi.org/10.1029/2020JD032759>, 2020.

Kelp, M. M., Jacob, D. J., Lin, H., and Sulprizio, M. P.: An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry, *J. Adv. Model. Earth Sy.*, 14(6), e2021MS002926, <https://doi.org/10.1029/2021MS002926>, 2022.

Kodros, J. K., and Pierce, J. R.: Important global and regional differences in aerosol cloud-albedo effect estimates between simulations with and without prognostic aerosol microphysics, *J. Geophys. Res. Atmos.*, 122(7), 4003–4018, <https://doi.org/10.1002/2016JD025886>, 2017.

Kinnison, D. E., Brasseur, G. P., and Walters, S.: Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model, *J. Geophys. Res.-Atmos.*, 112, D15, <https://doi.org/10.1029/2003JD004473>, 2007.

Lamarque, J.-F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J., and Tyndall, G. K.: CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model, *Geosci. Model Dev.*, 5, 369–411, <https://doi.org/10.5194/gmd-5-369-2012>, 2012.

- 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, *P. Natl. Acad. Sci. USA*, 118, 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, 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 Br: 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, H.: jimmielin/HEMCO3-Paper-Code: Release code for HEMCO 3.0 paper (Version rel2), Zenodo [code], <https://doi.org/10.5281/zenodo.4706173>, 2021.
- 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.
- Lin, H., Jacob, D. J., Lundgren, E. W., Sulprizio, M. P., Keller, C. A., Fritz, T. M., Eastham, S. D., Emmons, L. K., Campbell, P. C., Baker, B., Saylor, R. D., and Montuoro, R.: 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, *Geosci. Model Dev.*, 14, 5487–5506, <https://doi.org/10.5194/gmd-14-5487-2021>, 2021.
- Lin, J.-T. and McElroy, M. B.: Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing, *Atmos. Environ.*, 44, 1726–1739, 2010.
- Lin, S.-J.: A “Vertically Lagrangian” Finite-Volume Dynamical Core for Global Models, *Mon. Weather Rev.*, 132, 2293–2307, [https://doi.org/10.1175/1520-0493\(2004\)132<2293:AVLFDC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2), 2004.
- Lin, S.-J. and Rood, R. B.: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes, *Mon. Weather. Rev.*, 124, 2046–2070, [https://doi.org/10.1175/1520-0493\(1996\)124<2046:MFFSLT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2046:MFFSLT>2.0.CO;2), 1996.
- 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, 1575–1593, [https://doi.org/10.1175/1520-0493\(1994\)122<1575:ACOTVL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<1575:ACOTVL>2.0.CO;2), 1994.
- Liu, H., Jacob, D. J., Bey, I., and Yantosca, R. M.: Constraints from ²¹⁰Pb and ⁷Be on wet deposition and transport in a global three-dimensional chemical tracer model driven by assimilated meteorological fields, *J. Geophys. Res.*, 106, 12109–12128, <https://doi.org/10.1029/2000JD900839>, 2001.

Liu, X., Ma, P.-L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., Ghan, S. J., and Rasch, P. J.: Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within version 5.3 of the Community Atmosphere Model, *Geosci. Model Dev.*, 9, 505–522, <https://doi.org/10.5194/gmd-9-505-2016>, 2016.

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-o2) as an atmospheric chemistry module for Earth system models, *Geosci. Model Dev.*, 8, 595–602, <https://doi.org/10.5194/gmd-8-595-2015>, 2015.

Long, M. S., and Lin, H.: KineticPreProcessor/KPP-AR-boxmodel: GC_v13.4-AR-BoxModel [Software], Zenodo, <https://doi.org/10.5281/zenodo.6791657>, 2022.

Lu, X., Zhang, L., Wu, T., Long, M. S., Wang, J., Jacob, D. J., Zhang, F., Zhang, J., Eastham, S. D., Hu, L., Zhu, L., Liu, X., and Wei, M.: Development of the global atmospheric chemistry general circulation model BCC-GEOS-Chem v1.0: model description and evaluation, *Geosci. Model Dev.*, 13, 3817–3838, <https://doi.org/10.5194/gmd-13-3817-2020>, 2020.

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.

Mahowald, N. M., Lamarque, J.-F., Tie, X. X., and Wolff, E.: Sea-salt aerosol response to climate change: Last Glacial Maximum, preindustrial, and doubled carbon dioxide climates, *J. Geophys. Res. Atmos.*, 111(D5), D05303, <https://doi.org/10.1029/2005JD006459>, 2006a.

Mahowald, N. M., Muhs, D. R., Levis, S., Rasch, P. J., Yoshioka, M., Zender, C. S., and Luo, C.: Change in atmospheric mineral aerosols in response to climate: Last glacial period, preindustrial, modern, and doubled carbon dioxide climates, *J. Geophys. Res. Atmos.*, 111(D10), D10202, <https://doi.org/10.1029/2005JD006653>, 2006b.

Mao, J., Jacob, D. J., Evans, M. J., Olson, J. R., Ren, X., Brune, W. H., Clair, J. M. St., Crounse, J. D., Spencer, K. M., Beaver, M. R., Wennberg, P. O., Cubison, M. J., Jimenez, J. L., Fried, A., Weibring, P., Walega, J. G., Hall, S. R., Weinheimer, A. J., Cohen, R. C., Chen, G., Crawford, J. H., McNaughton, C., Clarke, A. D., Jaeglé, L., Fisher, J. A., Yantosca, R. M., Le Sager, P., and Carouge, C.: Chemistry of hydrogen oxide radicals (HOx) in the Arctic troposphere in spring, *Atmos. Chem. Phys.*, 10, 5823–5838, <https://doi.org/10.5194/acp-10-5823-2010>, 2010.

Marais, E. A. and Wiedinmyer, C.: Air Quality Impact of Diffuse and Inefficient Combustion Emissions in Africa (DICE-Africa), *Environ. Sci. Technol.*, 2016, 50, 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.

McDuffie, E. E., Martin, R. V., Spadaro, J. V., Burnett, R., Smith, S. J., O'Rourke, P., Hammer, M. S., van Donkelaar, A., Bindle, L., Shah, V., Jaeglé, L., Luo, G., Yu, F., Adeniran, J. A., Lin, J.,

and Brauer, M.: Source sector and fuel contributions to ambient PM_{2.5} and attributable mortality across multiple spatial scales, *Nat. Commun.*, 12, 3594, <https://doi.org/10.1038/s41467-021-23853-y>, 2021.

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.

Moch, J. M., Dovrou, E., Mickley, L. J., Keutsch, F. N., Liu, Z., Wang, Y., Dombek, T. L., Kuwata, M., Budisulistiorini, S. H., Yang, L., Decesari, S., Paglione, M., Alexander, B., Shao, J., Munger, J. W., and Jacob, D. J.: Global Importance of Hydroxymethanesulfonate in Ambient Particulate Matter: Implications for Air Quality, *J. Geophys. Res.-Atmos.*, 125, e2020JD032706, <https://doi.org/10.1029/2020jd032706>, 2020.

Moch, J. M., Mickley, L. J., Keller, C. A., Bian, H., Lundgren, E. W., Zhai, S., and Jacob, D. J.: Aerosol-Radiation Interactions in China in Winter: Competing Effects of Reduced Shortwave Radiation and Cloud-Snowfall-Albedo Feedbacks Under Rapidly Changing Emissions, *J. Geophys. Res. Atmos.*, 127(9), e2021JD035442, <https://doi.org/10.1029/2021JD035442>, 2022.

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., Mills, G. E., Stevenson, D. S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., and Williams, M. L.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, *Atmos. Chem. Phys.*, 15, 8889–8973, <https://doi.org/10.5194/acp-15-8889-2015>, 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.

Naik, V., Voulgarakis, A., Fiore, A. M., Horowitz, L. W., Lamarque, J.-F., Lin, M., Prather, M. J., Young, P. J., Bergmann, D., Cameron-Smith, P. J., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R., Eyring, V., Faluvegi, G., Folberth, G. A., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., van Noije, T. P. C., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R., Shindell, D. T., Stevenson, D. S., Stroe, S., Sudo, K., Szopa, S., and Zeng, G.: Preindustrial to present-day changes in tropospheric hydroxyl radical and methane lifetime from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), *Atmos. Chem. Phys.*, 13, 5277–5298, <https://doi.org/10.5194/acp-13-5277-2013>, 2013.

Neu, J. L. and Prather, M. J.: Toward a more physical representation of precipitation scavenging in global chemistry models: cloud overlap and ice physics and their impact on tropospheric ozone, *Atmos. Chem. Phys.*, 12, 3289–3310, <https://doi.org/10.5194/acp-12-3289-2012>, 2012.

Olivier, J., Peters, J., Granier, C., Petron, G., Muller, J. F., and Wallens, S.: Present and future surface emissions of atmospheric compounds, POET report #2, EU project EVK2-1999-00011, available at: http://accent.aero.jussieu.fr/Documents/del2_final.doc (lastaccess: 23July2021), 2003.

- 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 NO_x and its vertical distribution calculated from three-dimensional cloud-scale chemical transport model simulations, *J. Geophys. Res.-Atmos.*, **115**, D04301, <https://doi.org/10.1029/2009JD011880>, 2010.
- Pai, S. J., Heald, C. L., Pierce, J. R., Farina, S. C., Marais, E. A., Jimenez, J. L., Campuzano-Jost, P., Nault, B. A., Middlebrook, A. M., Coe, H., Shilling, J. E., Bahreini, R., Dingle, J. H., and Vu, K.: An evaluation of global organic aerosol schemes using airborne observations, *Atmos. Chem. Phys.*, **20**, 2637–2665, <https://doi.org/10.5194/acp-20-2637-2020>, 2020.
- Park, R., 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. Atmos.*, **109**(D15), 2003JD004474, <https://doi.org/10.1029/2003JD004474>, 2004.
- Park, R. J., Oak, Y. J., Emmons, L. K., Kim, C.-H., Pfister, G. G., Carmichael, G. R., Saide, P. E., Cho, S.-Y., Kim, S., Woo, J.-H., Crawford, J. H., Gaubert, B., Lee, H.-J., Park, S.-Y., Jo, Y.-J., Gao, M., Tang, B., Stanier, C. O., Shin, S. S., Park, H. Y., Bae, C., and Kim, E.: Multi-model inter-comparisons of air quality simulations for the KORUS-AQ campaign, *Elem. Sci. Anth.*, **9**, 00139, <https://doi.org/10.1525/elementa.2021.00139>, 2021.
- 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), *B. Am. Meteorol. Soc.*, **101**, E1743–E1760, <https://doi.org/10.1175/BAMS-D-19-0331.1>, 2020.
- Philip, S., Martin, R. V., and Keller, C. A.: Sensitivity of chemistry-transport model simulations to the duration of chemical and transport operators: a case study with GEOS-Chem v10-01, *Geosci. Model Dev.*, **9**, 1683–1695, <https://doi.org/10.5194/gmd-9-1683-2016>, 2016.
- 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**, 044018, <https://doi.org/10.1088/1748-9326/aa65a4>, 2017.
- Pound, R. J., Sherwen, T., Helmig, D., Carpenter, L. J., and Evans, M. J.: Influences of oceanic ozone deposition on tropospheric photochemistry, *Atmos. Chem. Phys.*, **20**, 4227–4239, <https://doi.org/10.5194/acp-20-4227-2020>, 2020.
- Prather, M. J., Holmes, C. D., Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, *Geophys. Res. Lett.*, **39**(9), L09803, <https://doi.org/10.1029/2012GL051440>, 2012.

- Price, C., Penner, J., and Prather, M.: NO_x from lightning: 1. Global distribution based on lightning physics, *J. Geophys. Res. Atmos.*, 102(D5), 5929–5941, <https://doi.org/10.1029/96JD03504>, 1997.
- Putman, W. M. and Lin, S.-J.: Finite-volume transport on various cubed-sphere grids, *J. Comput. Phys.*, 227, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>, 2007.
- Pye, H. O. T., Chan, A. W. H., Barkley, M. P., and Seinfeld, J. H.: Global modeling of organic aerosol: the importance of reactive nitrogen (NO_x and NO₃), *Atmos. Chem. Phys.*, 10, 11261–11276, <https://doi.org/10.5194/acp-10-11261-2010>, 2010.
- Ramboll Environment and Health: Comprehensive Air Quality Model with Extensions (CAMx) User's Guide, Version 7.10, available at: https://camx-wp.azurewebsites.net/Files/CAMxUsersGuide_v7.10.pdf (last access: 2 September 2021), 2020.
- Randerson, J. T., van der Werf, G. R., Giglio, L., Collatz, G. J., and Kasibhatla, P. S.: Global Fire Emissions Database, Version 4, (GFEDv4), ORNL DAAC, Oak Ridge, Tennessee, USA, <https://doi.org/10.3334/ORNLDAAC/1293>, 2018.
- Rastigejev, Y., Brenner, M. P., and Jacob, D. J.: Spatial reduction algorithm for atmospheric chemical transport models, *P. Natl. A. Sci.*, 104(35), 13875–13880, <https://doi.org/10.1073/pnas.0705649104>, 2007.
- 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.
- Randles, C. A., da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinzuka, Y., and Flynn, C. J.: The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation, *J. Climate*, 30, 6823–6850, <https://doi.org/10.1175/JCLI-D-16-0609.1>, 2017.
- Rémy, S., Veira, A., Paugam, R., Sofiev, M., Kaiser, J. W., Marengo, 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, 27, 92 pp., 2008.
- Rosenbrock, H. H.: Some general implicit processes for the numerical solution of differential equations, *The Computer Journal*, 5(4), 329–330, <https://doi.org/10.1093/comjnl/5.4.329>, 1963.
- Sander, R., Baumgaertner, A., Cabrera-Perez, D., Frank, F., Gromov, S., Grooß, J.-U., Harder, H., Huijnen, V., Jöckel, P., Karydis, V. A., Niemeyer, K. E., Pozzer, A., Riede, H., Schultz, M. G., Taraborrelli, D., and Tauer, S.: The community atmospheric chemistry box model CAABA/MECCA-4.0, *Geosci. Model Dev.*, 12, 1365–1385, <https://doi.org/10.5194/gmd-12-1365-2019>, 2019.
- Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, *Atmos. Chem. Phys.*, 6, 187–195, <https://doi.org/10.5194/acp-6-187-2006>, 2006.

194/acp-6-187-2006, 2006.

Sandu, A., Sander, R., Long, M. S., Yantosca, R. M., Lin, H., Shen, L., and Jacob, D. J.: Kinetic Pre-Processor/KPP: The Kinetic Pre-Processor (KPP) 3.0.0 [Software], Zenodo, <https://doi.org/10.5281/zenodo.7308373>, 2022.

Sandu, A., Verwer, J., Blom, J., Spee, E., Carmichael, G., and Potra, F.: Benchmarking stiff ode solvers for atmospheric chemistry problems II: Rosenbrock solvers, *Atmos. Environ.*, 31(20), 3459–3472, [https://doi.org/10.1016/S1352-2310\(97\)83212-8](https://doi.org/10.1016/S1352-2310(97)83212-8), 1997.

Santillana, M., Le Sager, P., Jacob, D. J., and Brenner, M. P.: An adaptive reduction algorithm for efficient chemical calculations in global atmospheric chemistry models, *Atmos. Environ.*, 44(35), 4426–4431, <https://doi.org/10.1016/j.atmosenv.2010.07.044>, 2010.

Schwantes, R. H., Emmons, L. K., Orlando, J. J., Barth, M. C., Tyndall, G. S., Hall, S. R., Ullmann, K., St. Clair, J. M., Blake, D. R., Wisthaler, A., and Bui, T. P. V.: Comprehensive isoprene and terpene gas-phase chemistry improves simulated surface ozone in the southeastern US, *Atmos. Chem. Phys.*, 20, 3739–3776, <https://doi.org/10.5194/acp-20-3739-2020>, 2020.

Schwantes, R. H., Lacey, F. G., Tilmes, S., Emmons, L. K., Lauritzen, P. H., Walters, S., Callaghan, P., Zarzycki, C. M., Barth, M. C., Jo, D. S., Bacmeister, J. T., Neale, R. B., Vitt, F., Kluzek, E., Roozitalab, B., Hall, S. R., Ullmann, K., Warneke, C., Peischl, J., Pollack, I. B., Flocke, F., Wolfe, G. M., Hanisco, T. F., Keutsch, F. N., Kaiser, J., Bui, T. P. V., Jimenez, J. L., Campuzano-Jost, P., Apel, E. C., Hornbrook, R. S., Hills, A. J., Yuan, B., and Wisthaler, A.: Evaluating the impact of chemical complexity and horizontal resolution on tropospheric ozone over the conterminous US with a global variable resolution chemistry model. *J. Adv. Model. Earth Sy.*, 14, e2021MS002889, <https://doi.org/10.1029/2021MS002889>, 2022.

Shah, V., Jacob, D. J., Dang, R., Lamsal, L. N., Strode, S. A., Steenrod, S. D., Boersma, K. F., Eastham, S. D., Fritz, T. M., Thompson, C., Peischl, J., Bourgeois, I., Pollack, I. B., Nault, B. A., Cohen, R. C., Campuzano-Jost, P., Jimenez, J. L., Andersen, S. T., Carpenter, L. J., Sherwen, T., and Evans, M. J.: Nitrogen oxides in the free troposphere: implications for tropospheric oxidants and the interpretation of satellite NO₂ measurements, *Atmos. Chem. Phys.*, 23, 1227–1257, <https://doi.org/10.5194/acp-23-1227-2023>, 2023.

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, *Remote Sens. Environ.*, 260, 112461, <https://doi.org/10.1016/j.rse.2021.112461>, 2021.

Shen, L., Jacob, D. J., Santillana, M., Wang, X., and Chen, W.: An adaptive method for speeding up the numerical integration of chemical mechanisms in atmospheric chemistry models: application to GEOS-Chem version 12.0.0, *Geosci. Model Dev.*, 13, 2475–2486, <https://doi.org/10.5194/gmd-13-2475-2020>, 2020.

Shen, L., Jacob, D. J., Santillana, M., Bates, K., Zhuang, J., and Chen, W.: A machine-learning-guided adaptive algorithm to reduce the computational cost of integrating kinetics in global atmospheric chemistry models: application to GEOS-Chem versions 12.0.0 and 12.9.1, *Geosci. Model Dev.*, 15, 1677–1687, <https://doi.org/10.5194/gmd-15-1677-2022>, 2022.

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 oxidants: a global model study in GEOS-Chem, *Atmos. Chem. Phys.*, 16, 1161–1186, <https://doi.org/10.5194/acp-16-1161-2016>, 2016a.

Sherwen, T., Schmidt, J. A., Evans, M. J., Carpenter, L. J., Großmann, K., Eastham, S. D., Jacob, D. J., Dix, B., Koenig, T. K., Sinreich, R., Ortega, I., Volkamer, R., Saiz-Lopez, A., Prados-Roman, C., Mahajan, A. S., and Ordóñez, C.: Global impacts of tropospheric halogens (Cl, Br, I) on oxidants and composition in GEOS-Chem, *Atmos. Chem. Phys.*, 16, 12239–12271, <https://doi.org/10.5194/acp-16-12239-2016>, 2016b.

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, OPENSky, <https://doi.org/10.5065/D6854MVH>, 2008.

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, 5415–5424, <https://doi.org/10.1016/j.atmosenv.2011.07.012>, 2011.

Suarez, M., Trayanov, A., Hill, C., Schopf, P., and Vikhliav, 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, <https://doi.org/10.1145/1297385.1297388>, 2007.

Tang, Z., Jiang, Z., Chen, J., Yang, P., and Shen, Y.: The capabilities of the adjoint of GEOS-Chem model to support HEMCO emission inventories and MERRA-2 meteorological data, *Geosci. Model Dev.*, 16, 6377–6392, <https://doi.org/10.5194/gmd-16-6377-2023>, 2023.

The International GEOS-Chem Community: jimmielin/geos-chem: AR-GC-archival_13.4 [Software], Zenodo, <https://doi.org/10.5281/zenodo.6791655>, 2022.

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, *B. Am. Meteorol. Soc.*, 97, 1229–1247, <https://doi.org/10.1175/BAMS-D-14-00164.1>, 2016.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Conley, A., Schultz, M. G., Saunio, M., Thouret, V., Thompson, A. M., Oltmans, S. J., Johnson, B., and Tarasick, D.: Technical Note: Ozone sonde climatology between 1995 and 2011: description, evaluation and applications, *Atmos. Chem. Phys.*, 12, 7475–7497, <https://doi.org/10.5194/acp-12-7475-2012>, 2012.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Ma, P.-L., Liu, X., Ghan, S., Bardeen, C., Arnold, S., Deeter, M., Vitt, F., Ryerson, T., Elkins, J. W., Moore, F., Spackman, J. R., and Val Martin, M.: Description and evaluation of tropospheric chemistry and aerosols in the Community Earth System Model (CESM1.2), *Geosci. Model Dev.*, 8, 1395–1426, <https://doi.org/10.5194/gmd-8-1395-2015>, 2015.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., Smith, A. K., Neely, R. R., Conley, A., Vitt, F., Val Martin, M., Tanimoto, H., Simpson, I., Blake, D. R., and

- Blake, N.: Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), *Geosci. Model Dev.*, 9, 1853–1890, <https://doi.org/10.5194/gmd-9-1853-2016>, 2016.
- Travis, K. R., Heald, C. L., Allen, H. M., Apel, E. C., Arnold, S. R., Blake, D. R., Brune, W. H., Chen, X., Commane, R., Crounse, J. D., Daube, B. C., Diskin, G. S., Elkins, J. W., Evans, M. J., Hall, S. R., Hints, E. J., Hornbrook, R. S., Kasibhatla, P. S., Kim, M. J., Luo, G., McKain, K., Millet, D. B., Moore, F. L., Peischl, J., Ryerson, T. B., Sherwen, T., Thames, A. B., Ullmann, K., Wang, X., Wennberg, P. O., Wolfe, G. M., and Yu, F.: Constraining remote oxidation capacity with ATom observations, *Atmos. Chem. Phys.*, 20, 7753–7781, <https://doi.org/10.5194/acp-20-7753-2020>, 2020.
- 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, 2493–2512, <https://doi.org/10.1002/2016JD025767>, 2017.
- United States Environmental Protection Agency: CMAQ (Version 5.2.1), Zenodo [code], <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.
- van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., Arneeth, 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, 193–199, <https://doi.org/10.1021/es0700799>, 2008.
- Wang, H., Easter, R. C., Rasch, P. J., Wang, M., Liu, X., Ghan, S. J., Qian, Y., Yoon, J.-H., Ma, P.-L., and Vinoj, V.: Sensitivity of remote aerosol distributions to representation of cloud–aerosol interactions in a global climate model, *Geosci. Model Dev.*, 6, 765–782, <https://doi.org/10.5194/gmd-6-765-2013>, 2013.
- Wang, X., Jacob, D. J., Downs, W., Zhai, S., Zhu, L., Shah, V., Holmes, C. D., Sherwen, T., Alexander, B., Evans, M. J., Eastham, S. D., Neuman, J. A., Veres, P. R., Koenig, T. K., Volkamer, R., Huey, L. G., Bannan, T. J., Percival, C. J., Lee, B. H., and Thornton, J. A.: Global tropospheric halogen (Cl, Br, I) chemistry and its impact on oxidants, *Atmos. Chem. Phys.*, 21, 13973–13996, <https://doi.org/10.5194/acp-21-13973-2021>, 2021.

Wells, K. C., Millet, D. B., Payne, V. H., Deventer, M. J., Bates, K. H., de Gouw, J. A., Graus, M., Warneke, C., Wisthaler, A., and Fuentes, J. D.: Satellite isoprene retrievals constrain emissions and atmospheric oxidation, *Nature*, 585, 225–233, <https://doi.org/10.1038/s41586-020-2664-3>, 2020.

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 NO_x, sea salt aerosols, and biogenic volatile organic compounds, *Sci. Data*, 7, 148, <https://doi.org/10.1038/s41597-020-0488-5>, 2020.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, *Earth Syst. Sci. Data*, 9, 697–720, <https://doi.org/10.5194/essd-9-697-2017>, 2017.

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.

Wofsy, S. C., Afshar, S., Allen, H. M., Apel, E., Asher, E. C., Barletta, B., Bent, J., Bian, H., Biggs, B. C., Blake, D. R., Blake, N., Bourgeois, I., Brock, C. A., Brune, W. H., Budney, J. W., Bui, T. P., Butler, A., Campuzano-Jost, P., Chang, C. S., Chin, M., Commane, R., Correa, G., Crounse, J. D., Cullis, P. D., Daube, B. C., Day, D. A., Dean-Day, J. M., Dibb, J. E., DiGangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Erdesz, F., Fiore, A. M., Flynn, C. M., Froyd, K., Gesler, D. W., Hall, S. R., Hanisco, T. F., Hannun, R. A., Hills, A. J., Hintsa, E. J., Hoffman, A., Hornbrook, R. S., Huey, L. G., Hughes, S., Jimenez, J. L., Johnson, B. J., Katich, J. M., Keeling, R. F., Kim, M. J., Kupc, A., Lait, L. R., Lamarque, J.-F., Liu, J., McKain, K., McLaughlin, R. J., Meinardi, S., Miller, D. O., Montzka, S. A., Moore, F. L., Morgan, E. J., Murphy, D. M., Murray, L. T., Nault, B. A., Neuman, J. A., Newman, P. A., Nicely, J. M., Pan, X., Paplawsky, W., Peischl, J., Prather, M. J., Price, D. J., Ray, E., Reeves, J. M., Richardson, M., Rollins, A. W., Rosenlof, K. H., Ryerson, T. B., Scheuer, E., Schill, G. P., Schroder, J. C., Schwarz, J. P., St. Clair, J. M., Steenrod, S. D., Stephens, B. B., Strode, S. A., Sweeney, C., Tanner, D., Teng, A. P., Thames, A. B., Thompson, C. R., Ullmann, K., Veres, P. R., Vieznor, N., Wagner, N. L., Watt, A., Weber, R., Weinzierl, B., Wennberg, P., Williamson, C. J., Wilson, J. C., Wolfe, G. M., Woods, C. T., and Zeng, L. H.: ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols. ORNL DAAC, Oak Ridge, Tennessee, USA, <https://doi.org/10.3334/ORNLDAAC/1581>, 2018.

Woo, J.-H., Kim, Y., Kim, H.-K., Choi, K.-C., Eum, J.-H., Lee, J.-B., Lim, J.-H., Kim, J., and Seong, M.: Development of the CREATE Inventory in Support of Integrated Climate and Air Quality Modeling for Asia, *Sustainability*, 12, 7930, <https://doi.org/10.3390/su12197930>, 2020.

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. Atmos.*, 112(D5), D05302, <https://doi.org/10.1029/2006JD007801>, 2007.

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, 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, 1876–1895, <https://doi.org/10.1016/j.rse.2010.03.012>, 2010.

Yang, L. H., Jacob, D. J., Colombi, N. K., Zhai, S., Bates, K. H., Shah, V., Beaudry, E., Yantosca, R. M., Lin, H., Brewer, J. F., Chong, H., Travis, K. R., Crawford, J. H., Lamsal, L. N., Koo, J.-H., and Kim, J.: Tropospheric NO₂ vertical profiles over South Korea and their relation to oxidant chemistry: implications for geostationary satellite retrievals and the observation of NO₂ diurnal variation from space, *Atmos. Chem. Phys.*, 23, 2465–2481, <https://doi.org/10.5194/acp-23-2465-2023>, 2023.

Young, T. R., and Boris, J. P.: A numerical technique for solving stiff ordinary differential equations associated with the chemical kinetics of reactive-flow problems, *J. Phys. Chem.*, 81(25), 2424–2427, <https://doi.org/10.1021/j100540a018>, 1977.

Young, P. J., Archibald, A. T., Bowman, K. W., Lamarque, J.-F., Naik, V., Stevenson, D. S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Horowitz, L. W., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R. B., Shindell, D. T., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), *Atmos. Chem. Phys.*, 13, 2063–2090, <https://doi.org/10.5194/acp-13-2063-2013>, 2013.

Young, P. J., Naik, V., Fiore, A. M., Gaudel, A., Guo, J., Lin, M. Y., Neu, J. L., Parrish, D. D., Rieder, H. E., Schnell, J. L., Tilmes, S., Wild, O., Zhang, L., Ziemke, J. R., Brandt, J., Delcloo, A., Doherty, R. M., Geels, C., Hegglin, M. I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M. G., Woodhouse, M. T., and Zeng, G.: Tropospheric Ozone Assessment Report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, *Elem. Sci. Anth.*, 6, 10, <https://doi.org/10.1525/elementa.265>, 2018.

Yu, F. and Luo, G.: Simulation of particle size distribution with a global aerosol model: contribution of nucleation to aerosol and CCN number concentrations, *Atmos. Chem. Phys.*, 9, 7691–7710, <https://doi.org/10.5194/acp-9-7691-2009>, 2009.

Yu, K., Keller, C. A., Jacob, D. J., Molod, A. M., Eastham, S. D., and Long, M. S.: Errors and improvements in the use of archived meteorological data for chemical transport modeling: an analysis using GEOS-Chem v11-01 driven by GEOS-5 meteorology, *Geosci. Model Dev.*, 11, 305–319, <https://doi.org/10.5194/gmd-11-305-2018>, 2018.

Zender, C. S., Bian, H., and Newman, D.: Mineral Dust Entrainment and Deposition (DEAD) model: Description and 1990s dust climatology, *J. Geophys. Res.-Atmos.*, 108, 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., and Rees, S. L.: Toward Convective-Scale Prediction within the Next Generation Global Prediction System, *B. Am. Meteorol. Soc.*, 100, 1225–1243, <https://doi.org/10.1175/BAMS-D-17-0246.1>, 2019.

Zhao, S., Russell, M. G., Hakami, A., Capps, S. L., Turner, M. D., Henze, D. K., Percell, P. B., Resler, J., Shen, H., Russell, A. G., Nenes, A., Pappin, A. J., Napelenok, S. L., Bash, J. O., Fahey, K. M., Carmichael, G. R., Stanier, C. O., and Chai, T.: A multiphase CMAQ version 5.0 adjoint, *Geosci. Model Dev.*, 13, 2925–2944, <https://doi.org/10.5194/gmd-13-2925-2020>, 2020.

Zhao, Y., Saunio, M., Bousquet, P., Lin, X., Berchet, A., Hegglin, M. I., Canadell, J. G., Jackson, R. B., Hauglustaine, D. A., Szopa, S., Stavert, A. R., Abraham, N. L., Archibald, A. T., Bekki, S., Deushi, M., Jöckel, P., Josse, B., Kinnison, D., Kirner, O., Marécal, V., O'Connor, F. M., Plummer, D. A., Revell, L. E., Rozanov, E., Stenke, A., Strode, S., Tilmes, S., Dlugokencky, E. J., and Zheng, B.: Inter-model comparison of global hydroxyl radical (OH) distributions and their impact on atmospheric methane over the 2000–2016 period, *Atmos. Chem. Phys.*, 19, 13701–13723, <https://doi.org/10.5194/acp-19-13701-2019>, 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.