

1 **The capabilities of the adjoint of GEOS-Chem model to support HEMCO**  
2 **emission inventories and MERRA-2 meteorological data**

3  
4 Zhaojun Tang<sup>1</sup>, Zhe Jiang<sup>1\*</sup>, Jiaqi Chen<sup>1</sup>, Panpan Yang<sup>1</sup>, Yanan Shen<sup>1</sup>

5  
6 <sup>1</sup>School of Earth and Space Sciences, University of Science and Technology of China, Hefei,  
7 Anhui, 230026, China.

8  
9 \*Correspondence to: Zhe Jiang ([zhejiang@ustc.edu.cn](mailto:zhejiang@ustc.edu.cn))

10  
11  
12 **Abstract**

13 Adjoint of the GEOS-Chem model has been widely used to constrain the sources of  
14 atmospheric compositions. Here we designed a new framework to facilitate emission inventory  
15 updates in the adjoint of GEOS-Chem model. The major advantage of this new framework is  
16 good readability and extensibility, which allows us to support Harmonized Emissions  
17 Component (HEMCO) emission inventories conveniently and to easily add more emission  
18 inventories following future updates in GEOS-Chem forward simulations. Furthermore, we  
19 developed new modules to support MERRA-2 meteorological data, which allows us to perform  
20 long-term analysis with consistent meteorological data in 1979-present. The performances of  
21 the developed capabilities were evaluated with the following steps: 1) diagnostic outputs of  
22 carbon monoxide (CO) sources and sinks to ensure the correct reading and use of emission  
23 inventories; 2) forward simulations to compare the modeled surface and column CO  
24 concentrations among various model versions; 3) backward simulations to compare adjoint  
25 gradients of global CO concentrations to CO emissions with finite difference gradients; and 4)  
26 observing system simulation experiments (OSSE) to evaluate the model performance in 4D  
27 variational (4D-var) assimilations. Finally, an example application of 4D-var assimilation was  
28 presented to constrain anthropogenic CO emissions in 2015 by assimilating Measurement of  
29 Pollution in the Troposphere (MOPITT) CO observations. The capabilities developed in this  
30 work are important for better applications of the adjoint of GEOS-Chem model in the future.

31 These capabilities will be submitted to the standard GEOS-Chem adjoint code base for better  
32 development of the community of the adjoint of GEOS-Chem model.

33

## 34 **1. Introduction**

35 GEOS-Chem is a global 3D chemical transport model (CTM) and has been widely used  
36 to analyze the sources and variabilities of atmospheric compositions (Whaley et al., 2015; Li  
37 et al., 2019; Hammer et al., 2020; Jiang et al., 2022). GEOS-Chem model is driven by  
38 meteorological reanalysis data from the Goddard Earth Observing System (GEOS) of the  
39 Global Modeling and Assimilation Office (GMAO). Emissions in GEOS-Chem model are  
40 calculated with state-of-the-art inventories such as CEDS (Community Emissions Data  
41 System) (Hoesly et al., 2018), MIX (Li et al., 2017) and NEI2011 (National Emissions  
42 Inventory). Based on GEOS-Chem forward simulation, the adjoint of the GEOS-Chem model  
43 (Henze et al., 2007) further provides the capability of backward simulation of physical and  
44 chemical processes within the 4D variational (4D-var) framework. The major advantage of the  
45 adjoint model is obtaining the sensitivity of atmospheric concentrations to multiple model  
46 variables within a single backward simulation. The major applications of the adjoint of GEOS-  
47 Chem model include inverse analyses of atmospheric composition emissions by minimizing  
48 the difference between simulations and observations (Jiang et al., 2015a; Zhang et al., 2018;  
49 Qu et al., 2022) as well as sensitivity analyses to analyze the sources of atmospheric  
50 compositions (Jiang et al., 2015b; Zhao et al., 2019; Dedoussi et al., 2020).

51 The algorithm of the 4D-var framework requires identical model processes in the forward  
52 and backward simulations. Ideally, the code for the adjoint model should be updated following  
53 the GEOS-Chem forward codes to take advantage of the new features in GEOS-Chem forward  
54 simulations. However, the updates in the adjoint model are difficult and usually delayed. For  
55 example, the MEERA-2 meteorological reanalysis data with temporal coverage of 1979-

56 present were supported in the GEOS-Chem forward simulations in v11-01. The adjoint of  
57 GEOS-Chem model does not support MERRA-2, and thus, long-term analysis must combine  
58 different meteorological reanalysis data, such as GEOS-4 (1985-2007), GEOS-5 (2004-2012)  
59 and GEOS-FP (2012-present). For instance, Jiang et al. (2017) constrained global carbon  
60 monoxide (CO) emissions in 2001-2015, while the derived trends in CO emissions in Jiang et  
61 al. (2017) could be affected by the discontinuity among various versions of the meteorological  
62 data (i.e., GEOS-4 in 2001-2003, GEOS-5 in 2004-2012 and GEOS-FP in 2013-2015) and the  
63 lack of consistency in the model physics of GEOS-5.

64 Emission inventories play a key role in the simulation of atmospheric compositions.  
65 Harmonized Emissions Component (HEMCO) (Keller et al., 2014; Lin et al., 2021) was  
66 included in the GEOS-Chem forward simulations in v10-01. HEMCO is responsible for inputs  
67 of meteorological and emission data with default support for emission inventories such as  
68 CEDS, MIX and NEI2011. New emission inventories can be added readily within HEMCO  
69 framework. There are noticeable differences between HEMCO and the adjoint of GEOS-Chem  
70 model. First, meteorological and emission data are read with individual modules in the adjoint  
71 of GEOS-Chem model. Second, the inputs of emission inventories are undertaken by different  
72 modules that were developed individually with significant discrepancies in the source code. In  
73 addition, the file format (e.g., binary punch in the adjoint of GEOS-Chem that is the format of  
74 older GEOS-Chem versions in contrast to netCDF in HEMCO), emission variables and the  
75 usage methods of emission variables (e.g., emission hierarchy, scaling factors and time slice)  
76 are inconsistent. These differences have posed a barrier to the application of new emission  
77 inventories in the adjoint of GEOS-Chem model.

78 The lack of support to the updated emission inventories can affect the applications of the  
79 adjoint of GEOS-Chem model. First, adjoint-based sensitivity analyses are obtained by the  
80 backward simulations of atmospheric compositions (i.e., adjoint tracers) and the combination

81 of adjoint tracers with emissions. Out-of-date emission inventories can thus result in inaccurate  
82 estimation of the adjoint sensitivities. Second, while inverse analyses are constrained by  
83 atmospheric observations, the updated emission inventories are still critical because they are  
84 helpful for better convergence of 4D-var assimilations by setting a more reasonable a priori  
85 penalty in the cost function. For instance, the a priori biomass burning CO emissions (GFED3,  
86 van der Werf et al. (2010)) in Jiang et al. (2017) lack interannual variabilities later than 2011.  
87 In order to obtain reasonable convergence of biomass burning emissions, the a priori biomass  
88 burning emissions in September-November 2006 were applied to September-November 2015  
89 over Indonesia in Jiang et al. (2017).

90 Ideally, people should consider porting the complete HEMCO to the adjoint of GEOS-  
91 Chem model to match the new features in GEOS-Chem forward simulations. However, a  
92 complete port of HEMCO implies replacing the input framework of the adjoint of GEOS-Chem  
93 model, as well as restructuring of HEMCO and the adjoint of GEOS-Chem model to address  
94 the compatibility issues, which is very challenging and may not be necessary because the  
95 meteorological modules still work well in the adjoint of GEOS-Chem model. Consequently, a  
96 major objective of this work is to design a new framework to facilitate emission inventory  
97 updates in the adjoint of GEOS-Chem model. For this objective, this new framework must have  
98 good readability and extensibility to allow us to support HEMCO emission inventories  
99 conveniently and to add more emissions inventories following future updates in GEOS-Chem  
100 forward simulations easily. Furthermore, we developed new modules to support MERRA-2  
101 meteorological data within the current framework of the adjoint of GEOS-Chem model, as  
102 reuse of existing frameworks can save much work.

103 CO is one of the most important atmospheric pollutants and plays a key role in  
104 tropospheric chemistry. Sources of atmospheric CO include fossil fuel combustion, biomass  
105 burning and oxidation of hydrocarbons. The major sink of atmospheric CO is hydroxyl

106 radical (OH). The simple chemical sink of atmospheric CO allows us to simulate atmospheric  
107 CO with linearized chemistry; for example, the tagged-CO mode of the GEOS-Chem model  
108 can reduce the calculation cost by 98% with respect to the full chemistry mode by reading  
109 archived monthly OH fields. The tagged-CO mode of the GEOS-Chem model has been widely  
110 used to investigate the sources and variabilities of atmospheric CO in recent decades (Heald et  
111 al., 2004; Kopacz et al., 2009; Jiang et al., 2017). The capabilities developed in this work are  
112 thus based on the tagged-CO mode, as it can effectively accelerate the model development  
113 process. More efforts are needed in the future to extend these capabilities to support emissions  
114 inventories associated with full chemistry simulations.

115 The results presented in this paper show the development, integration, evaluation, and  
116 application of these new capabilities, which is important to better applications of the adjoint of  
117 GEOS-Chem model in the future. The capabilities developed in this work will be submitted to  
118 the standard GEOS-Chem adjoint code base (Henze et al., 2007) for better development of the  
119 community of the adjoint of GEOS-Chem model. This paper is organized as follows: in Section  
120 2, we describe the adjoint of GEOS-Chem model, the development of these new capabilities,  
121 and the Measurement of Pollution in the Troposphere (MOPITT) CO observations used in this  
122 work. In Section 3, we evaluated the performances of the developed capabilities in forward and  
123 backward simulations, together with observing system simulation experiments (OSSE) to  
124 evaluate the model performance in 4D-var assimilations. An example application of 4D-var  
125 assimilation to constrain anthropogenic CO emissions in 2015 by assimilating MOPITT CO  
126 observations was also presented. Our conclusions follow in Section 4.

127

## 128 **2. Methodology and Data**

### 129 **2.1 Adjoint of the GEOS-Chem model**

130 We use version v35n of the adjoint of GEOS-Chem model. Our analysis is conducted at

131 a horizontal resolution of  $4^\circ \times 5^\circ$  with 47 vertical levels and employs the CO-only simulation  
 132 (tagged-CO mode). The global default anthropogenic emission inventory in the standard  
 133 version of the adjoint of GEOS-Chem model (hereafter referred to as GC-Adjoint-STD) is  
 134 Global Emissions Initiative (GEIA), but is replaced by the following regional emission  
 135 inventories: NEI2008 in North America, the Criteria Air Contaminants (CAC) inventory for  
 136 Canada, the Big Bend Regional Aerosol and Visibility Observational (BRAVO) Study  
 137 Emissions Inventory for Mexico (Kuhns et al., 2003), the Cooperative Program for Monitoring  
 138 and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP) inventory  
 139 for Europe in 2000 (Vestreng and Klein, 2002) and the INTEX-B Asia emissions inventory for  
 140 2006 (Zhang et al., 2009). Biomass burning emissions are based on the GFED3 (van der Werf  
 141 et al., 2010).

142 The objective of the 4D-var approach is to minimize the difference between simulations  
 143 and observations described by the cost function (Henze et al., 2007):

$$144 \quad J(\mathbf{x}) = \sum_{i=1}^N (\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i)^T \mathbf{S}_\Sigma^{-1} (\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i) + \gamma (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) \quad (1)$$

145 where  $\mathbf{x}$  is the state vector of CO emissions,  $N$  is the number of observations that are  
 146 distributed in time over the assimilation period,  $\mathbf{z}_i$  is a given measurement, and  $\mathbf{F}(\mathbf{x})$  is the  
 147 forward model. The error estimates are assumed to be Gaussian and are given by  $\mathbf{S}_\Sigma$ , the  
 148 observational error covariance matrix, and  $\mathbf{S}_a$ , the a priori error covariance matrix. The cost  
 149 function is minimized through minimizing the adjoint gradients by adjusting the CO emissions  
 150 iteratively:

$$151 \quad \nabla_{\mathbf{x}} J(\mathbf{x}) = \sum_{k=1}^N \left[ 2\mathbf{S}_\Sigma^{-1} (\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i) \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}} \right] + 2\gamma \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) \quad (2)$$

152 We assume a uniform observation error of 20%. The combustion CO sources (fossil fuel,  
 153 biofuel and biomass burning) and the oxidation source from biogenic volatile organic  
 154 compounds (VOCs) are combined, assuming a 50% uniform a priori error. We optimize the  
 155 source of CO from the oxidation of methane ( $\text{CH}_4$ ) separately as an aggregated global source,

156 assuming an a priori uncertainty of 25%. The CO emission estimates are optimized with  
157 monthly temporal resolution. Following Jiang et al. (2017), we performed 40 iterations  
158 (forward + backward simulations) for each month, which usually produced 6-8 accepted  
159 iterations (i.e., successful line searches in the large-scale bound constrained optimization (L-  
160 BFGS-B, Zhu et al. (1997)) to reduce the cost functions and adjoint gradients. The a posteriori  
161 CO emission estimates were calculated based on the last accepted iteration, which usually  
162 corresponded to the iteration with the lowest cost function.

## 163 **2.2 New framework to read emission inventories**

164 A major objective of this work is to design a new framework to facilitate emission  
165 inventory updates in the adjoint of GEOS-Chem model. As shown in Fig. 1, we first initialize  
166 the array in [INITIAL] and batch read the emission data in [READ\_DATA], which were  
167 interpolated offline with  $1^\circ \times 1^\circ$  resolution by considering the mass conservation. Here, the data  
168 include the emission inventory data listed in Table S1 (see the SI), the corresponding scaling  
169 factor data and the mask map files of domain definitions. The data are scaled in  
170 [SCALE\_DATA] by multiplying the corresponding annual, season, month, week, and 24-hour  
171 emission factors and are then online interpolated to the current resolution ( $4^\circ \times 5^\circ$  in this work)  
172 of the model by [RGRID\_DATA], which was followed by the application of region masks in  
173 [MASK].

174 The emission variable of CO obtained in this part is written to the model memory in  
175 emission.f and emission\_adj.f by calling DO\_EMISSIONS to ensure the consistent emissions  
176 in both forward and backward simulations. The GET\_[TRACER] subroutines are used to  
177 obtain the CO emission variable, which participates in the calculation of physicochemical  
178 processes in the model, to interact with other modules. Finally, the variable is cleaned from the  
179 memory by the [CLEANUP] module. It should be noted that a two-step interpolation is  
180 employed in this work (hereafter referred to as GC-Adjoint-HEMCO) following GC-Adjoint-

181 STD, for example,  $0.1^\circ \times 0.1^\circ$  to  $1^\circ \times 1^\circ$  and then to  $4^\circ \times 5^\circ$  for the NEI2011 inventory, which is  
182 different from the one-step interpolation in GEOS-Chem forward model (v12-08-01, hereafter  
183 referred to as GC-v12), for example,  $0.1^\circ \times 0.1^\circ$  to  $4^\circ \times 5^\circ$  directly for the NEI2011 inventory.  
184 The different interpolation methods can lead to differences in the interpolated emission data.

### 185 **2.3 Updates in emission inventories**

186 In addition to baseline emission data, there are critical factors that affect the usage of  
187 emission data in the models. Reading the emission data correctly thus does not necessarily  
188 mean using emission data correctly. For example, emission hierarchy is used to prioritize  
189 emission fields within the same emission category. Emissions of higher hierarchy overwrite  
190 lower hierarchy data. Regional emission inventories usually have a higher hierarchy within  
191 their mask boundaries. Scaling factors are used to adjust the baseline emissions with annual,  
192 season, month, week, and 24-hour temporal scales. Time slice selection is used to define the  
193 usage methods of the emission data outside the original temporal range; for instance, data can  
194 be interpreted as climatology and recycled once the end of the last time slice is reached or be  
195 only considered as long as the simulation time is within the time range. Consequently, we must  
196 validate the integrated emissions carefully to ensure that the abovementioned factors have been  
197 correctly applied and to ensure that the calculated emissions are reasonable for individual  
198 inventories and the combination of all inventories.

199 To take advantage of this new framework, six HEMCO emission inventories have been  
200 added to this work. To validate the emissions, we performed actual simulations with GC-v12,  
201 GC-Adjoint-HEMCO and GC-Adjoint-STD, and the emissions were calculated in the model  
202 simulations and then output to the Log file. As shown in Table S1, the CEDS emission  
203 inventory ( $0.5^\circ \times 0.5^\circ$ ) is adopted in GC-Adjoint-HEMCO to provide global default emissions  
204 for 1750-2019. The diurnal scale factors are applied to obtain CO emissions at different  
205 moments of the day. Fig. S1 (see the SI) shows CEDS CO emissions in 2015 in GC-v12 and



206 GC-Adjoint-HEMCO and GEIA CO emissions in GC-Adjoint-STD, and we find noticeable  
207 differences in CO emissions between CEDS and GEIA. As shown in Table 1, the CEDS CO  
208 emissions in 2015 were 613.57 and 613.85 Tg/y in GC-v12 and GC-Adjoint-HEMCO,  
209 respectively, with a relative difference of 0.05% between GC-v12 and GC-Adjoint-HEMCO.  
210 The GEIA CO emissions in 2015 were 445.88 Tg/year in GC-Adjoint-STD.

211 The default CEDS inventory is replaced by the following regional emission inventories  
212 in GC-Adjoint-HEMCO: MIX in Asia ( $0.25^{\circ} \times 0.25^{\circ}$ ), NEI2011 in the United States  
213 ( $0.1^{\circ} \times 0.1^{\circ}$ ), DICE\_AFRICA and EDGARV43 in Africa ( $0.1^{\circ} \times 0.1^{\circ}$ ) and APEI in Canada  
214 ( $0.1^{\circ} \times 0.1^{\circ}$ ). As shown in Fig. S2 (see the SI), the MIX inventory provides Asian emissions in  
215 2008-2010, accompanied by diurnal scale factors to describe daily emission variation. The  
216  $1^{\circ} \times 1^{\circ}$  scale factors in the AnnualScalar.geos.1x1.nc file further provide the annual variation in  
217 1985-2010. As shown in Table 1, the MIX CO emissions in 2015 were 321.18 and 321.71 Tg/y  
218 in GC-v12 and GC-Adjoint-HEMCO, respectively, with a relative difference of 0.17% between  
219 GC-v12 and GC-Adjoint-HEMCO. The INTEX-B CO emissions in 2015 were 353.03 Tg/y in  
220 GC-Adjoint-STD.

221 The NEI2011 inventory (Fig. S3, see the SI) provides anthropogenic emissions for the  
222 United States in 2011 with annual scalar factors from 2006-2013. The weekday and weekend  
223 factors are read from NEI99.dow.geos.1x1.nc file since 1999 with all CO factors of 1.0 on  
224 weekdays and between 0.990 and 0.997 on Saturdays and Sundays. The NEI2011 CO  
225 emissions in 2015 were 35.83 and 37.70 Tg/y in GC-v12 and GC-Adjoint-HEMCO,  
226 respectively, with a relative difference of 5.22% between GC-v12 and GC-Adjoint-HEMCO.  
227 The NEI2008 CO emissions in 2015 were 52.87 Tg/y in GC-Adjoint-STD. APEI (Fig. S4, see  
228 the SI) is the primary source of anthropogenic emissions in the Canadian domain. The APEI  
229 CO emissions in 2015 were 6.10 and 6.17 Tg/y in GC-v12 and GC-Adjoint-HEMCO,  
230 respectively, with a relative difference of 1.14% between GC-v12 and GC-Adjoint-HEMCO.

231 The CAC CO emissions in 2015 were 10.20 Tg/y in GC-Adjoint-STD. Following GC-v12, the  
232 CO emissions in APEI are enhanced by 19% to account for coemitted VOC in the tagged-CO  
233 simulation.

234 Emissions for the African domain are provided by the combination of DICE\_AFRICA  
235 and EDGARV43 (Fig. S5, see the SI). Here DICE\_AFRICA includes anthropogenic and  
236 biofuel emissions in 2013. We read the DICE\_AFRICA emissions data into the model in two  
237 types according to the guidelines of the inventory. Emissions from sectors such as automobiles  
238 and motorcycles are aggregated into anthropogenic sources, and household-generated  
239 emissions such as charcoal and agricultural waste are aggregated into biofuel sources. Efficient  
240 combustion emissions from EDGAR v4.3 in 1970-2010 then compensate for the lacking  
241 sources in DICE\_AFRICA. Daily variation factors for CO are also used here for emissions  
242 across the African region. The 2010 CO seasonal scale factors are used in EDGAR v4.3 for  
243 sectoral emission sources. The DICE\_AFRICA and EDGARV43 CO emissions in 2015 were  
244 83.42 and 83.02 Tg/y in GC-v12 and GC-Adjoint-HEMCO, respectively, with a relative  
245 difference of -0.48% between GC-v12 and GC-Adjoint-HEMCO. Following GC-v12, the CO  
246 emissions in DICE\_AFRICA and EDGARV43 are enhanced by 19% to account for coemitted  
247 VOC in the tagged-CO simulation.

248 The biomass burning emission inventory in GC-Adjoint-HEMCO is GFED4 (Fig. S6,  
249 see the SI), which includes dry matter emissions from a total of seven sectors in 1997-2019.  
250 The same GFED\_emssion\_factors.H header file as in the GC-v12 version is read in the GC-  
251 Adjoint-HEMCO. This file contains the ratio factors of atmospheric pollutants, and we  
252 multiply the ratio factors one by one according to the ID of each species to ensure that the  
253 species in the model have biomass burning sources. The GFED4 CO emissions in 2015 were  
254 437.13 and 435.89 Tg/y in GC-v12 and GC-Adjoint-HEMCO, respectively, with a relative  
255 difference of -0.28% between GC-v12 and GC-Adjoint-HEMCO. The GFED3 CO emissions

256 in 2015 were 382.04 Tg/year in GC-Adjoint-STD. Following GC-v12, the combustion CO  
257 sources in biomass burning are enhanced by 5% to consider the CO generated by VOC in the  
258 tagged-CO simulation.

259 Fig. 2 shows the total combustion CO emissions in 2015 from GC-v12, GC-Adjoint-  
260 HEMCO and GC-Adjoint-STD. As shown in Table 2, the regional combustion CO emissions  
261 are 320.66 and 320.38 Tg/y (Asia), 73.96 and 66.93 Tg/y (North America), 199.51 and  
262 193.29/y Tg (Africa), 79.04 and 78.91 Tg/y (South America), 31.58 and 30.96 Tg/y (Europe)  
263 and 12.24 and 11.99 Tg/y (Australia) in GC-v12 and GC-Adjoint-HEMCO, respectively. Fig.  
264 3 further shows the monthly combustion CO emissions in 2015 from GC-v12, GC-Adjoint-  
265 HEMCO and GC-Adjoint-STD, and there are good agreements in the monthly variation of CO  
266 emissions between GC-v12 and GC-Adjoint-HEMCO. The CO emissions in GC-Adjoint-STD  
267 are similar to those in GC-v12 and GC-Adjoint-HEMCO in winter and spring but with large  
268 differences in summer and autumn. This seasonal difference may reflect the influence of  
269 different emission inventories on biomass burning.

#### 270 **2.4 Updates in CO chemical sources and sinks**

271 The biogenic emissions in GC-Adjoint-STD are Model of Emissions of Gases and  
272 Aerosols from Nature, version 2.0 (MEGANv2.0, Guenther et al. (2006)) in the full chemistry  
273 simulation but are GEIA in the tagged-CO simulation (Fig. S7, see the SI). Fisher et al. (2017)  
274 demonstrated improvement in modeled CO concentrations in tagged-CO simulation by reading  
275 archived VOC- and CH<sub>4</sub>-generated CO fields provided by full chemistry simulation. The  
276 archived VOC- and CH<sub>4</sub>-generated CO fields in 2013 (PCO\_3Dglobal.geosfp.4x5.nc) were set  
277 as the default CO chemical sources in the tagged-CO simulation in GC-v12 and supported in  
278 GC-Adjoint-HEMCO. As shown in Table 2, the CO chemical sources (columns) obtained by  
279 reading the archived VOC- and CH<sub>4</sub>-generated CO fields demonstrate good agreement between  
280 GC-v12 and GC-Adjoint-HEMCO. However, they are 30-60% lower than those in GEIA in

281 GC-Adjoint-STD, and this difference could be partially associated with the inconsistency  
282 between the archived VOC-generated CO fields in 2013 and the actual meteorological data in  
283 2015 in the simulation.

284 The default CH<sub>4</sub>-generated CO emissions in GC-Adjoint-STD (Fig. S8, see the SI) are  
285 calculated based on averaged CH<sub>4</sub> concentrations in four latitude bands (90°S - 30°S, 30°S -  
286 00°S, 00°N - 30°N, 30°N - 90°N), which are based on Climate Monitoring and Diagnostics  
287 Laboratory (CMDL) surface observations and Intergovernmental Panel on Climate Change  
288 (IPCC) future scenarios. As shown in Table 2, there are good agreements in the CH<sub>4</sub>-generated  
289 CO emissions between GC-v12 and GC-Adjoint-HEMCO by reading  
290 PCO\_3Dglobal.geosfp.4x5.nc, and they are 20-60% lower than those in CMDL/IPCC in GC-  
291 Adjoint-STD. Furthermore, the default archived monthly OH fields were updated following  
292 GC-v12 with updated calculations for the decay rate (KRATE, from JPL 03 to JPL 2006) in  
293 GC-Adjoint-HEMCO. The subsequent CO sinks (Fig. S9, see the SI) in GC-v12 and GC-  
294 Adjoint-HEMCO are 20-40% higher than those in GC-Adjoint-STD.

## 295 **2.5 Updates in meteorological data**

296 The MERRA-2 meteorological data (1979-present) are supported in GC-Adjoint-  
297 HEMCO to ensure long-term consistency in the meteorological data in the analyses. The code  
298 porting to support MERRA-2 follows the current framework of the adjoint of GEOS-Chem  
299 model, particularly because the meteorological variables and vertical resolutions of MERRA-  
300 2 are the same as those of GEOS-FP (2012-present), while GEOS-FP is already supported by  
301 GC-Adjoint-STD. Fig. 4A-B show the averages of surface CO concentrations in 2015 from  
302 GC-Adjoint-HEMCO driven by MERRA-2 and GEOS-FP, respectively. Our results  
303 demonstrate lower surface CO concentrations driven by MERRA-2 (Fig. 4C), although there  
304 is good agreement in the spatial distributions of CO concentrations. Similarly, Fig. 4D-F show  
305 the averages of CO columns in 2015 from GC-Adjoint-HEMCO driven by MERRA-2 and

306 GEOS-FP and their differences. Despite the noticeable differences in surface CO  
307 concentrations (Fig. 4C), the differences in CO columns (Fig. 4F) are much smaller, and the  
308 modeled CO columns driven by MERRA-2 are higher than those driven by GEOS-FP over the  
309 Indian Ocean. The discrepancy between surface and column CO in Fig. 4 may reflect the  
310 impacts of different convective transports on the modeled CO concentrations.

## 311 **2.6 MOPITT CO measurements**

312 The MOPITT data used here were obtained from the joint retrieval (V7J) of CO from  
313 thermal infrared (TIR, 4.7 $\mu$ m) and near-infrared (NIR, 2.3 $\mu$ m) radiances using an optimal  
314 estimation approach (Worden et al., 2010; Deeter et al., 2017). The retrieved volume mixing  
315 ratios (VMR) are reported as layer averages of 10 pressure levels with a footprint of 22 km  $\times$   
316 22 km. Following Jiang et al. (2017), we reject MOPITT data with CO column amounts less  
317 than  $5 \times 10^{17}$  molec/cm<sup>2</sup> and with low cloud observations. Since the NIR channel measures  
318 reflected solar radiation, only daytime data are considered.

319

## 320 **3. Model evaluation and application**

### 321 **3.1 Model performances in forward and backward simulations**

322 The reasonable emissions in the diagnostic outputs in Section 2 do not necessarily mean  
323 the correct integration of emissions in the assimilations. Consequently, here we evaluate the  
324 performance of GC-Adjoint-HEMCO in forward simulations. Fig. 5 shows the averages of  
325 surface and column CO concentrations in 2015 from GC-v12, GC-Adjoint-HEMCO and GC-  
326 Adjoint-STD. As shown in Table 2, the regional differences between GC-v12 and GC-Adjoint-  
327 HEMCO are 2.6%, -5.7%, -4.6%, -1.7%, -1.4% and -3.6% in surface CO concentrations, and  
328 -2.3%, -3.6%, -3.3%, -3.1%, -3.3% and -4.1% in CO columns over Asia, North America,  
329 Africa, South America, Europe, and Australia, respectively. There are larger regional  
330 differences in CO concentrations between GC-v12 and GC-Adjoint-STD: 4.6%, -10.1%, 6.3%,

331 22.5%, 6.4% and 25.7% in surface CO concentrations, and -0.7%, -9.9%, 2.5%, 8.0%, -5.8%  
332 and 8.5% in CO columns over Asia, North America, Africa, South America, Europe, and  
333 Australia, respectively. The agreement between GC-v12 and GC-Adjoint-HEMCO confirms  
334 the reliability of GC-Adjoint-HEMCO in forward simulations, while the small differences in  
335 CO concentrations between GC-v12 and GC-Adjoint-HEMCO are expected in view of the  
336 comparable differences in regional emissions, chemical sources and sinks, as shown in Table  
337 2.

338 In addition to forward simulations, the reliability of 4D-var assimilation also relies on  
339 the accuracy of the adjoint-based sensitivities, which are obtained by the backward simulations  
340 of adjoint tracers and the combination of adjoint tracers with emissions. As mentioned in  
341 Section 2.2, we have made corresponding modifications to both forward and backward  
342 modules. Consequently, here we further evaluate the performance of GC-Adjoint-HEMCO in  
343 backward simulations. Here the adjoint gradients are simplified as:

$$344 \quad \nabla_x J(\mathbf{x}) = \frac{\partial F_N}{\partial \mathbf{x}} \quad (3)$$

345 The adjoint gradients (Eq. 3) represent the sensitivities of modeled atmospheric compositions  
346 at the final time step (i.e.,  $i = N$ ) to emissions, which were then compared with the finite  
347 difference gradients calculated with:

$$348 \quad \Lambda = \frac{J(\mathbf{x}+\delta\mathbf{x})-J(\mathbf{x}-\delta\mathbf{x})}{2\delta\mathbf{x}} \quad (4)$$

349 Here the finite difference gradients represent the response of modeled atmospheric  
350 compositions at the final time step to finite perturbations in emissions provided by the forward  
351 simulations ( $\delta\mathbf{x} = 10\%$  in this work).

352 Fig. 6A-C show the comparison of adjoint and finite difference gradients of global  
353 surface CO concentrations to CO emissions with a 24-hour assimilation window by turning on  
354 the convection, planetary boundary layer mixing and advection processes individually. We find  
355 good consistency in the gradients with respect to convection and planetary boundary layer

356 (PBL) mixing. The larger deviation with respect to advection is caused by the discrete  
357 advection algorithm in forward simulations and continuous advection algorithm in backward  
358 simulations (Henze et al., 2007). Fig. 6D-F further exhibit the effects of combined model  
359 processes (turning off advection as suggested by Henze et al. (2007)). We find good agreement  
360 between the adjoint and finite difference gradients with different assimilation windows (24  
361 hours, 7 days and one month). Moreover, Fig. S10 and S11 (see the SI) demonstrate the  
362 comparisons of sensitivities at higher model levels within the PBL and free troposphere by  
363 showing consistent results to Fig. 6. This confirms the consistency in the impacts of emissions  
364 to modeled atmospheric compositions between the forward and backward simulations, which  
365 is the prerequisite for more detailed evaluations in the following Sections.

### 366 **3.2 Observing system simulation experiments with pseudo-CO observations**

367 Here we further evaluate the performance of GC-Adjoint-HEMCO in 4D-var  
368 assimilations. OSSE is a useful method and has been widely used to evaluate the performance  
369 of various data assimilation systems (Jones et al., 2003; Barré et al., 2015; Shu et al., 2022). In  
370 contrast to assimilations by assimilating actual atmospheric observations, pseudo-observations  
371 are usually generated by model simulations and then assimilated in OSSE. The true  
372 atmospheric states are known in OSSEs as they are used to produce the pseudo-observations,  
373 and consequently, the difference between assimilated and true atmospheric states describes the  
374 capability of the assimilation systems to converge to the true atmospheric states in assimilations  
375 when assimilating actual observations.

376 The pseudo-observations in this work are produced by archiving CO concentrations from  
377 GC-Adjoint-HEMCO forward simulations with the CO emissions unchanged (i.e., the default  
378 CO emission inventory such as CEDS, MIX and NEI2011). According to the usage of pseudo-  
379 observations, two types of OSSE are performed in this work: 1) full modeled CO fields are  
380 assimilated as pseudo-observations so that we have pseudo-CO observations at every grid/level

381 and time step (hereafter referred to as OSSE-FullOBS). This experiment is designed to evaluate  
382 the performance of the assimilation system under ideal conditions with full coverage of  
383 observations. 2) The modeled CO fields are sampled at the locations/times of MOPITT CO  
384 observations and smoothed with MOPITT a priori concentrations and averaging kernels to  
385 produce MOPITT-like pseudo-CO observations (hereafter referred to as OSSE-MOPITT). This  
386 experiment is designed to evaluate the performance of the assimilation system under actual  
387 conditions with limited coverage of observations.

388 In the inverse analysis with the pseudo-CO observations, we reduce the anthropogenic  
389 CO emissions by 50% so that the objective of the OSSE is to produce scaling factors that can  
390 return the source estimate to the default emissions (i.e., scaling factors of 1.0). Fig. 7A shows  
391 the annual scaling factors in 2015 in OSSE-FullOBS. After 40 iterations, the a posteriori  
392 anthropogenic CO emission estimates converge to the true states in all major emission regions.  
393 As shown in Table 3, the regional scaling factors of OSSE-FullOBS are 1.00, 0.97, 0.97, 1.00,  
394 0.98 and 0.94 for anthropogenic CO emissions over Asia, North America, Africa, South  
395 America, Europe, and Australia, respectively.

396 Furthermore, Fig. 7D shows the annual scaling factors in OSSE-MOPITT, which are  
397 noticeably worse than those in Fig. 7A. The regional scaling factors of OSSE-MOPITT are  
398 1.04, 0.88, 1.01, 1.02, 0.84 and 0.81 for anthropogenic CO emissions over Asia, North  
399 America, Africa, South America, Europe, and Australia, respectively. With respect to OSSE-  
400 FullOBS, the limited coverage of observations in OSSE-MOPITT has resulted in  
401 approximately 15% underestimations in the a posteriori CO emission estimates over North  
402 America and Europe. In addition, Fig. 7B-C and Fig. 7E-F show the a priori and a posteriori  
403 biases in the modeled CO columns. We find dramatic improvements in the modeled CO  
404 columns, which confirms the reliability of the 4D-var assimilation system. The difference  
405 between Fig. 7B and 6E reflects the influence of the application of MOPITT averaging kernels,



406 which lead to larger negative biases in the a priori simulation. It should be noted that we cannot  
407 expect comparable improvement in the actual assimilations because of the potential effects of  
408 model and observation errors.

### 409 **3.3 Anthropogenic CO emissions constrained with MOPITT CO observations**

410 As an example of the application of GC-Adjoint-HEMCO, here we constrain  
411 anthropogenic CO emissions in 2015 by assimilating MOPITT CO observations. Fig.8A shows  
412 the relative differences between modeled and MOPITT CO columns at the beginning of each  
413 month in 2015 (i.e., biases in monthly initial CO conditions) in the original GEOS-Chem  
414 simulations. We find dramatic underestimations in the modeled CO columns by approximately  
415 30-40%. As indicated by previous studies (Jiang et al., 2013; Jiang et al., 2017), the biases in  
416 monthly initial CO conditions are caused by model biases in CO concentrations accumulated  
417 in previous months. Considering that the lifetime of CO is approximately 2-3 months, the  
418 negative biases in the initial conditions can result in negative biases in the modeled CO  
419 concentration in the following month. A lack of consideration of these biases, as shown in Fig.  
420 8A, can thus result in overestimations in the derived monthly CO emission estimates because  
421 the assimilation system will tend to adjust emissions to reduce the initial condition-induced  
422 biases.

423 Following Jiang et al. (2017), a suboptimal sequential Kalman filter (Todling and Cohn,  
424 1994; Tang et al., 2022) was employed in this work to optimize the modeled CO concentrations  
425 with an hourly resolution by combining GC-Adjoint-HEMCO forward simulation and  
426 MOPITT CO observations. The CO concentrations provided by the Kalman filter assimilations  
427 were archived at the beginning of each month, which were used as the optimized monthly initial  
428 CO conditions in the inverse analysis. As shown in Fig. 8B, the biases in the modeled CO  
429 columns in the optimized initial CO conditions are pronounced lower than those in the original  
430 simulation (Fig. 8A). The optimization of the initial CO conditions is essential for our inverse

431 analysis, as it can ensure that the adjustments in CO emissions are dominated by the differences  
432 between simulations and observations in the current month instead of the 30-40%  
433 underestimations in CO columns accumulated in previous months.

434 Fig. 9A shows the distribution of a priori anthropogenic CO emissions in 2015. The  
435 regional a priori anthropogenic CO emissions (as shown in Table 4) are 243.53, 34.42, 23.24,  
436 30.39, 25.94 and 2.02 Tg/y over Asia, North America, Africa, South America, Europe, and  
437 Australia, respectively. As shown in Fig. 9B, our inverse analysis suggests a wide distribution  
438 of underestimations in the a priori anthropogenic CO emissions in 2015 except in E. China.  
439 The regional scaling factors (Table 4) are 1.16, 1.47, 1.52, 1.41, 1.60 and 1.38, and the a  
440 posteriori anthropogenic CO emissions are 283.20, 50.47, 35.34, 42.92, 41.62 and 2.79 Tg/y  
441 over Asia, North America, Africa, South America, Europe, and Australia, respectively. As  
442 shown in Fig. 9C, we find noticeable underestimations in the modeled CO columns in the a  
443 priori simulations, despite the negative biases being much weaker than those in Fig. 8A due to  
444 the optimization of the initial CO conditions. The negative biases are effectively reduced in the  
445 a posteriori simulation driven by the a posteriori CO emission estimates (Fig. 9D).

446 Finally, we compare the a posteriori CO emission estimates in this work with Jiang et al.  
447 (2017), who constrained CO emissions in 2001-2015 with GC-Adjoint-STD by assimilating  
448 the same MOPITT CO observations. As shown in Table 4, the a posteriori anthropogenic CO  
449 emission estimates in this work match well with Jiang et al. (2017) in North America and Africa  
450 but are 38%, 157% and 228% higher than those in Jiang et al. (2017) in Asia, South America  
451 and Australia, respectively. A major discrepancy between this work and Jiang et al. (2017) is  
452 the treatment of ocean grids. Jiang et al. (2017) defined ocean grids as continental boundary  
453 conditions, which were rewritten hourly using the optimized CO concentrations archived from  
454 the suboptimal sequential Kalman filter by assimilating MOPITT CO observations. Only  
455 MOPITT data over land were assimilated in the 4D-var assimilations in Jiang et al. (2017)

456 without any change in CO distribution over the ocean. In addition, the large differences in  
457 chemical sources and sinks between GC-Adjoint-HEMCO and GC-Adjoint-STD, for example,  
458 lower VOC-generated CO emissions by 40-60% and higher CO sinks by 20-40% in GC-  
459 Adjoint-HEMCO, as shown in Table 2, may also contribute to the discrepancy in the derived  
460 a posteriori CO emission estimates.

461 As shown in Fig. 9D, the a posteriori simulation demonstrates positive biases in CO  
462 columns over China and Southeast Asia, which is a signal of overestimated local CO emissions;  
463 meanwhile, the negative biases over the northern Pacific Ocean are reduced in the a posteriori  
464 simulation. The negative biases over the remote ocean are more affected by CO chemical  
465 sources and sinks; however, biases in chemical sources cannot be effectively adjusted because  
466 of the global uniform scaling factor for CH<sub>4</sub>-generated CO emissions; biases in chemical sinks  
467 cannot be adjusted because of the fixed OH fields in the tagged-CO simulation. Jiang et al.  
468 (2017) tried to address this problem by defining continental boundary conditions so that the  
469 inverse analysis is dominated by local MOPITT observations to avoid the influence of model  
470 biases accumulated within the long-range transport. Conversely, CO emissions over China and  
471 Southeast Asia are overestimated in this work to offset the negative biases over the northern  
472 Pacific Ocean. We expect similar overestimations in the a posteriori CO emission estimates  
473 over South America, southern Africa, and Australia in this work because it is the effective  
474 pathway to reduce the negative bias over the ocean in the Southern Hemisphere.

#### 475 **4. Conclusion**

476 This work demonstrates our efforts on the development of a new framework to facilitate  
477 emission inventory updates in the adjoint of GEOS-Chem model. The major advantage of this  
478 new framework is good readability and extensibility, which allows us to conveniently support  
479 HEMCO emission inventories, including CEDS, MIX, NEI2011, DICE\_AF, AF\_EDGAR43,  
480 APEI and GFED4. The updated emission inventories are critical for reliable sensitivity

481 analyses, as well as better convergence of assimilations by setting a more reasonable a priori  
482 penalty in the cost function. Second, we developed new modules to support MERRA-2  
483 meteorological data, which allows us to perform long-term inverse analysis with consistent  
484 meteorological data in 1979-present. We evaluated the performances of the developed  
485 capabilities by validating the diagnostic outputs of CO emissions, modeled surface and column  
486 CO concentrations in forward simulations, and adjoint gradients of global CO concentrations  
487 to CO emissions with respect to the finite difference gradients.

488 Two types of OSSE were conducted to evaluate the model performance in 4D-var  
489 assimilations. The a posteriori CO emissions converged to the true states in all major emission  
490 regions with fully covered pseudo-CO observations; the limited coverage of observations by  
491 sampling the pseudo-CO observations at the locations/times of MOPITT CO observations and  
492 smoothing with MOPITT averaging kernels resulted in approximately 15% underestimations  
493 in the a posteriori CO emissions over North America and Europe. Furthermore, as an example  
494 application of the developed capabilities, we constrain anthropogenic CO emissions in 2015  
495 by assimilating MOPITT CO observations. The a posteriori anthropogenic CO emission  
496 estimates derived in this work match well with Jiang et al. (2017) in North America and Africa  
497 but are overestimated in Asia, South America and Australia, which could be associated with  
498 the different treatment of MOPITT CO observations over ocean grids and the large differences  
499 in CO chemical sources and sinks. The capabilities developed in this work are a useful  
500 extension for the adjoint of GEOS-Chem model. More efforts are needed to support emissions  
501 inventories associated with full chemistry simulations, as well as integration of these  
502 capabilities with the standard GEOS-Chem adjoint code base for better development of the  
503 community of the adjoint of GEOS-Chem model.

504

505 **Code and data availability:** The MOPITT CO data can be downloaded from  
506 <https://asdc.larc.nasa.gov/data/MOPITT/>. The GEOS-Chem model (version 12.8.1) can be  
507 downloaded from [http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem\\_12#12.8.1](http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_12#12.8.1).  
508 The adjoint of GEOS-Chem model (GC-Adjoint-STD) can be downloaded from  
509 [http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem\\_Adjoint](http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint). The adjoint of  
510 GEOS-Chem model (GC-Adjoint-HEMCO) can be downloaded from  
511 <https://doi.org/10.5281/zenodo.7512111>.

512

513 **Author Contributions:** Z.J. designed the research. Z.T. developed the model code and  
514 performed the research. Z.J. and Z.T. wrote the manuscript. All authors contributed to  
515 discussions and editing the manuscript.

516

517 **Competing interests:** The authors declare that they have no conflicts of interest.

518

519 **Acknowledgments:** We thank the providers of the MOPITT CO data. The numerical  
520 calculations in this paper have been done on the supercomputing system in the Supercomputing  
521 Center of University of Science and Technology of China. This work was supported by the  
522 Hundred Talents Program of Chinese Academy of Science and National Natural Science  
523 Foundation of China (42277082, 41721002).

524

## 525 **Tables and Figures**

526 **Table 1.** CO emissions for each inventory in 2015 with unit Tg/y.

527

528 **Table 2.** Regional combustion CO emissions, VOC-generated CO (PCO\_NMVOC), CH<sub>4</sub>-  
529 generated CO (PCO\_CH<sub>4</sub>), CO sinks (CO\_OH, calculated as CO\_OH = KRATE×CO×OH),  
530 and simulated surface and column CO concentrations in 2015. The region definitions are shown  
531 in Fig. 2A.

532

533 **Table 3.** Annual scaling factors of anthropogenic CO emissions in OSSEs. The scaling factors  
534 represent the ratio of the estimated to true emissions. The ratio for the first guess is 0.5. The  
535 actual value is 1.0. The pseudo-observations are produced by GC-Adjoint-HEMCO forward  
536 simulation. The full modeled CO fields are used in OSSE-FullOBS as pseudo-CO observations.  
537 The modeled CO fields are smoothed with MOPITT averaging kernels to produce MOPITT-  
538 like pseudo-CO observations in OSSE-MOPITT.

539

540 **Table 4.** Regional anthropogenic CO emissions (with unit Tg/y) and annual scaling factors in  
541 2015 in this work and Jiang et al. 2017.

542

543 **Fig. 1.** Framework to read the updated emission inventories in GC-Adjoint-HEMCO.

544

545 **Fig. 2.** Total combustion CO emissions in 2015 from (a) GC-v12; (b) GC-Adjoint-HEMCO;  
546 (c) GC-Adjoint-STD. The unit is molec/cm<sup>2</sup>/s.

547

548 **Fig. 3.** Monthly variation in combustion CO emissions in 2015 from GC-v12, GC-Adjoint-  
549 HEMCO and GC-Adjoint-STD.

550

551 **Fig. 4.** Averages of surface CO concentrations (unit ppbv) in 2015 from (a) GC-Adjoint-  
552 HEMCO driven by MERRA-2, (b) GC-Adjoint-HEMCO driven by GEOS-FP and (c) their  
553 difference; (d-f) same as panels a-c, but for CO columns (column-averaged dry-air mole  
554 fractions, X<sub>co</sub>).

555

556 **Fig. 5.** Averages of surface CO concentrations (unit ppbv) in 2015 from (a) GC-v12; (b) GC-  
557 Adjoint-HEMCO; (c) GC-Adjoint-STD; (d-f) same as panels a-c, but for CO columns (column-  
558 averaged dry-air mole fractions, X<sub>co</sub>).

559

560 **Fig. 6.** Comparison of sensitivities of global CO concentrations (LFD\_GLOB and model level  
561 1) to CO emission scaling factors calculated using the adjoint method vs. the finite difference  
562 method. (a-c) the effects of convection, PBL mixing and advection with 24-hour assimilation  
563 window; (d-f) the combined effects (the advection process is turned off) with increased  
564 assimilation windows.

565

566 **Fig. 7.** (a) Annual scaling factors in OSSE-FullOBS. The scaling factors represent the ratio of  
567 the estimated to true emissions. The ratio for the first guess is 0.5. The actual value is 1.0. (b-  
568 c) the a priori and a posteriori biases calculated by (model-observation)/observation in OSSE-  
569 Full. (d-f) same as panels a-c, but for OSSE-MOPITT.

570

571 **Fig. 8.** (a) Biases in monthly initial CO conditions in 2015 in the original GEOS-Chem  
572 simulation. (b) same as panel a, but with optimized initial CO conditions provided by  
573 suboptimal sequential Kalman filter. The biases are calculated by (model-MOPITT)/MOPITT.

574

575 **Fig. 9.** (a) A priori anthropogenic CO emissions in 2015 with unit molec/cm<sup>2</sup>/s; (b) Annual  
576 scaling factors for CO emissions in 2015. The scaling factors represent the ratio of the estimated  
577 to true emissions. (c-d) the a priori and a posteriori biases calculated by (model-  
578 MOPITT)/MOPITT.

579

## 580 **References**

581 Barré, J., Edwards, D., Worden, H., Da Silva, A., and Lahoz, W.: On the feasibility of  
582 monitoring carbon monoxide in the lower troposphere from a constellation of Northern  
583 Hemisphere geostationary satellites. (Part 1), *Atmos Environ*, 113, 63-77,  
584 10.1016/j.atmosenv.2015.04.069, 2015.

585 Dedoussi, I. C., Eastham, S. D., Monier, E., and Barrett, S. R. H.: Premature mortality related  
586 to United States cross-state air pollution, *Nature*, 578, 261-265, 10.1038/s41586-020-1983-8,  
587 2020.

588 Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martínez-Alonso, S., Worden, H.  
589 M., and Sweeney, C.: A climate-scale satellite record for carbon monoxide: the MOPITT  
590 Version 7 product, *Atmos Meas Tech*, 10, 2533-2555, 10.5194/amt-10-2533-2017, 2017.

591 Fisher, J. A., Murray, L. T., Jones, D. B. A., and Deutscher, N. M.: Improved method for linear  
592 carbon monoxide simulation and source attribution in atmospheric chemistry models  
593 illustrated using GEOS-Chem v9, *Geosci Model Dev*, 10, 4129-4144, 10.5194/gmd-10-4129-  
594 2017, 2017.

595 Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of  
596 global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and  
597 Aerosols from Nature), *Atmos Chem Phys*, 6, 3181-3210, 10.5194/acp-6-3181-2006, 2006.

598 Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R.  
599 C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., Brauer, M., Apte, J. S., Henze, D. K.,  
600 Zhang, L., Zhang, Q., Ford, B., Pierce, J. R., and Martin, R. V.: Global Estimates and Long-  
601 Term Trends of Fine Particulate Matter Concentrations (1998-2018), *Environ Sci Technol*,  
602 54, 7879-7890, 10.1021/acs.est.0c01764, 2020.

603 Heald, C. L., Jacob, D. J., Jones, D. B. A., Palmer, P. I., Logan, J. A., Streets, D. G., Sachse,  
604 G. W., Gille, J. C., Hoffman, R. N., and Nehr Korn, T.: Comparative inverse analysis of  
605 satellite (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon  
606 monoxide, *J Geophys Res-Atmos*, 109, D23306, 10.1029/2004jd005185, 2004.

607 Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem,  
608 *Atmos Chem Phys*, 7, 2413-2433, 10.5194/acp-7-2413-2007, 2007.

609 Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T.,  
610 Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N.,  
611 Kurokawa, J.-i., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.:  
612 Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the  
613 Community Emissions Data System (CEDS), *Geosci Model Dev*, 11, 369-408, 10.5194/gmd-  
614 11-369-2018, 2018.

615 Jiang, Z., Jones, D. B. A., Worden, H. M., Deeter, M. N., Henze, D. K., Worden, J., Bowman,  
616 K. W., Brenninkmeijer, C. A. M., and Schuck, T. J.: Impact of model errors in convective  
617 transport on CO source estimates inferred from MOPITT CO retrievals, *J Geophys Res-*  
618 *Atmos*, 118, 2073-2083, 10.1002/jgrd.50216, 2013.

619 Jiang, Z., Jones, D. B. A., Worden, J., Worden, H. M., Henze, D. K., and Wang, Y. X.: Regional  
620 data assimilation of multi-spectral MOPITT observations of CO over North America, *Atmos*  
621 *Chem Phys*, 15, 6801-6814, 10.5194/acp-15-6801-2015, 2015a.

622 Jiang, Z., Worden, J. R., Jones, D. B. A., Lin, J. T., Verstraeten, W. W., and Henze, D. K.:  
623 Constraints on Asian ozone using Aura TES, OMI and Terra MOPITT, *Atmos Chem Phys*,  
624 15, 99-112, 10.5194/acp-15-99-2015, 2015b.

625 Jiang, Z., Worden, J. R., Worden, H., Deeter, M., Jones, D. B. A., Arellano, A. F., and Henze,  
626 D. K.: A 15-year record of CO emissions constrained by MOPITT CO observations, *Atmos*  
627 *Chem Phys*, 17, 4565-4583, 10.5194/acp-17-4565-2017, 2017.

628 Jiang, Z., Zhu, R., Miyazaki, K., McDonald, B. C., Klimont, Z., Zheng, B., Boersma, K. F.,  
629 Zhang, Q., Worden, H., Worden, J. R., Henze, D. K., Jones, D. B. A., Denier van der Gon,  
630 H. A. C., and Eskes, H.: Decadal Variabilities in Tropospheric Nitrogen Oxides Over United  
631 States, Europe, and China, *J Geophys Res-Atmos*, 127, e2021JD035872,  
632 10.1029/2021jd035872, 2022.

633 Jones, D. B. A., Bowman, K. W., Palmer, P. I., Worden, J. R., Jacob, D. J., Hoffman, R. N.,  
634 Bey, I., and Yantosca, R. M.: Potential of observations from the Tropospheric Emission  
635 Spectrometer to constrain continental sources of carbon monoxide, *J Geophys Res-Atmos*,  
636 108, 2003JD003702, 10.1029/2003jd003702, 2003.

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

640 Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q.: Comparison  
641 of adjoint and analytical Bayesian inversion methods for constraining Asian sources of carbon  
642 monoxide using satellite (MOPITT) measurements of CO columns, *Journal of Geophysical*  
643 *Research*, 114, D04305, 10.1029/2007jd009264, 2009.



644 Kuhns, H., Green, M., and Etyemezian, V.: Big Bend Regional Aerosol and Visibility  
645 Observational (BRAVO) Study Emissions Inventory, Report prepared for BRAVO Steering  
646 Committee, Desert Research Institute, Las Vegas, Nevada, 2003.

647 Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., and Bates, K. H.: Anthropogenic drivers of  
648 2013-2017 trends in summer surface ozone in China, *Proc Natl Acad Sci USA*, 116, 422-427,  
649 10.1073/pnas.1812168116, 2019.

650 Li, M., Zhang, Q., Kurokawa, J.-i., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D.  
651 G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and  
652 Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international  
653 collaboration framework of the MICS-Asia and HTAP, *Atmos Chem Phys*, 17, 935-963,  
654 10.5194/acp-17-935-2017, 2017.

655 Lin, H., Jacob, D. J., Lundgren, E. W., Sulprizio, M. P., Keller, C. A., Fritz, T. M., Eastham,  
656 S. D., Emmons, L. K., Campbell, P. C., Baker, B., Saylor, R. D., and Montuoro, R.:  
657 Harmonized Emissions Component (HEMCO) 3.0 as a versatile emissions component for  
658 atmospheric models: application in the GEOS-Chem, NASA GEOS, WRF-GC, CESM2,  
659 NOAA GEFS-Aerosol, and NOAA UFS models, *Geosci Model Dev*, 14, 5487-5506,  
660 10.5194/gmd-14-5487-2021, 2021.

661 Qu, Z., Henze, D. K., Worden, H. M., Jiang, Z., Gaubert, B., Theys, N., and Wang, W.:  
662 Sector - Based Top - Down Estimates of NO<sub>x</sub>, SO<sub>2</sub>, and CO Emissions in East Asia, *Geophys*  
663 *Res Lett*, 49, e2021GL096009, 10.1029/2021gl096009, 2022.

664 Shu, L., Zhu, L., Bak, J., Zoogman, P., Han, H., Long, X., Bai, B., Liu, S., Wang, D., Sun, W.,  
665 Pu, D., Chen, Y., Li, X., Sun, S., Li, J., Zuo, X., Yang, X., and Fu, T.-M.: Improved ozone  
666 simulation in East Asia via assimilating observations from the first geostationary air-quality  
667 monitoring satellite: Insights from an Observing System Simulation Experiment, *Atmos*  
668 *Environ*, 274, 119003, 10.1016/j.atmosenv.2022.119003, 2022.

669 Tang, Z., Chen, J., and Jiang, Z.: Discrepancy in assimilated atmospheric CO over East Asia  
670 in 2015–2020 by assimilating satellite and surface CO measurements, *Atmos Chem Phys*, 22,  
671 7815-7826, 10.5194/acp-22-7815-2022, 2022.

672 Todling, R., and Cohn, S. E.: Suboptimal schemes for atmospheric data assimilation based on  
673 the Kalman filter, *Monthly Weather Review*, 122, 10.1175/1520-  
674 0493(1994)122<2530:SSFADA>2.0.CO;2, 1994.

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

679 Vestreng, V., and Klein, H.: Emission data reported to UNECE/EMEP. Quality assurance and  
680 trend analysis and Presentation of WebDab, Norwegian Meteorological Institute, Oslo,  
681 Norway, 2002.

682 Whaley, C. H., Strong, K., Jones, D. B. A., Walker, T. W., Jiang, Z., Henze, D. K., Cooke, M.  
683 A., McLinden, C. A., Mittermeier, R. L., Pommier, M., and Fogal, P. F.: Toronto area ozone:  
684 Long-term measurements and modeled sources of poor air quality events, *J Geophys Res-*  
685 *Atmos*, 120, 11368-11390, 10.1002/2014JD022984, 2015.

686 Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P.:  
687 Observations of near-surface carbon monoxide from space using MOPITT multispectral  
688 retrievals, *Journal of Geophysical Research*, 115, D18314, 10.1029/2010jd014242, 2010.

689 Zhang, L., Chen, Y., Zhao, Y., Henze, D. K., Zhu, L., Song, Y., Paulot, F., Liu, X., Pan, Y.,  
690 Lin, Y., and Huang, B.: Agricultural ammonia emissions in China: reconciling bottom-up and  
691 top-down estimates, *Atmos Chem Phys*, 18, 339-355, 10.5194/acp-18-339-2018, 2018.

692 Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z.,  
693 Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.:  
694 Asian emissions in 2006 for the NASA INTEX-B mission, *Atmos Chem Phys*, 9, 5131-5153,  
695 10.5194/acp-9-5131-2009, 2009.

696 Zhao, H., Geng, G., Zhang, Q., Davis, S. J., Li, X., Liu, Y., Peng, L., Li, M., Zheng, B., Huo,  
697 H., Zhang, L., Henze, D. K., Mi, Z., Liu, Z., Guan, D., and He, K.: Inequality of household  
698 consumption and air pollution-related deaths in China, *Nat Commun*, 10, 4337,  
699 10.1038/s41467-019-12254-x, 2019.

700 Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines  
701 for Large-Scale Bound Constrained Optimization, *ACM Transactions on Mathematical*  
702 *Software*, 23, 550-560, 10.1145/279232.279236, 1997.

703

Inventories	GC-v12	GC-Adjoint-HEMCO	Inventories	GC-Adjoint-STD
CEDS	613.57	613.85	GEIA	445.88
MIX	321.18	321.71	INTEX-B	353.03
NEI2011	35.83	37.70	NEI2008	52.87
DICE_AF + AF_EDGAR43	83.42	83.02	\	\
APEI	6.10	6.17	CAC	10.20
GFED4	437.13	435.89	GFED3	382.04

**Table 1.** CO emissions for each inventory in 2015 with unit Tg/y.

Version Region	Combustion Emission (Tg/y)			PCO_NMVOC (kg/s)			PCO_CH4 (kg/s)		
	GC-v12	GC-Adjoint-HEMCO	GC-Adjoint-STD	GC-v12	GC-Adjoint-HEMCO	GC-Adjoint-STD	GC-v12	GC-Adjoint-HEMCO	GC-Adjoint-STD
Asia	320.66	320.38	331.65	15.49	15.52	22.37	14.21	14.40	10.67
North America	73.96	66.93	60.65	7.05	6.83	14.75	7.45	7.66	5.23
Africa	199.51	193.29	179.22	34.57	33.92	52.38	19.57	19.85	16.18
South America	79.04	78.91	75.82	44.15	42.55	74.64	17.14	17.42	14.08
Europe	31.58	30.96	48.48	4.20	4.14	10.17	7.13	7.41	4.58
Australia	12.24	11.99	22.87	21.23	20.68	48.89	13.88	14.62	10.67
Version Region	CO_OH (kg/s)			CO (surface ppbv)			CO (column xco)		
	GC-v12	GC-Adjoint-HEMCO	GC-Adjoint-STD	GC-v12	GC-Adjoint-HEMCO	GC-Adjoint-STD	GC-v12	GC-Adjoint-HEMCO	GC-Adjoint-STD
Asia	52.26	51.34	40.87	179.56	184.29	187.90	90.23	88.16	89.58
North America	23.02	22.57	16.20	120.38	113.49	108.27	79.16	76.27	71.35
Africa	63.78	61.84	51.03	133.56	127.38	141.97	84.26	81.52	86.36
South America	49.06	48.85	41.25	107.98	106.16	132.24	72.93	70.67	78.75
Europe	20.65	20.92	14.27	112.88	111.33	120.09	74.83	72.34	70.45
Australia	31.42	31.98	25.27	67.45	65.00	84.80	56.35	54.02	61.15

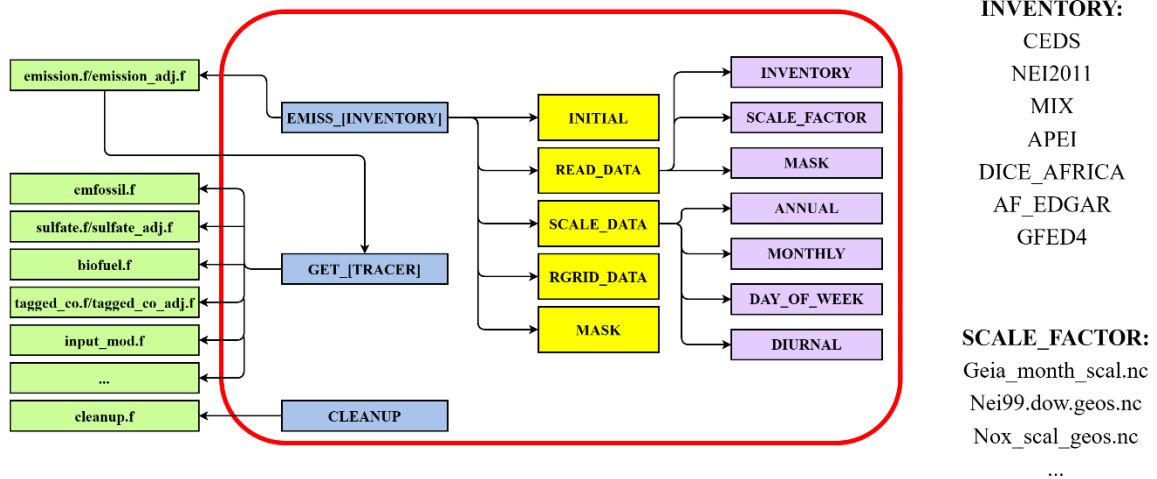
**Table 2.** Regional combustion CO emissions, VOC-generated CO (PCO\_NMVOC), CH<sub>4</sub>-generated CO (PCO\_CH<sub>4</sub>), CO sinks (CO\_OH, calculated as CO\_OH = KRATE×CO×OH), and simulated surface and column CO concentrations in 2015. The region definitions are shown in Fig. 2A.

	<b>Scaling Factors OSSE-FullOBS</b>	<b>Scaling Factors OSSE-MOPITT</b>
<b>Asia</b>	<b>1.00</b>	<b>1.04</b>
<b>North America</b>	<b>0.97</b>	<b>0.88</b>
<b>Africa</b>	<b>0.97</b>	<b>1.01</b>
<b>South America</b>	<b>1.00</b>	<b>1.02</b>
<b>Europe</b>	<b>0.98</b>	<b>0.84</b>
<b>Australia</b>	<b>0.94</b>	<b>0.81</b>

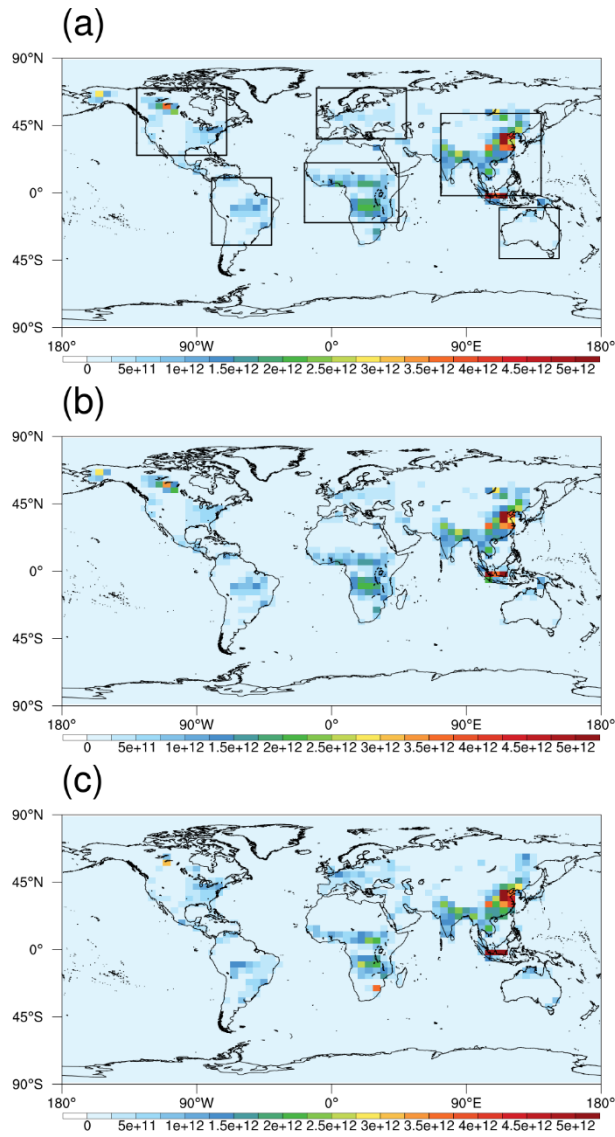
**Table 3.** Annual scaling factors of anthropogenic CO emissions in OSSEs. The scaling factors represent the ratio of the estimated to true emissions. The ratio for the first guess is 0.5. The actual value is 1.0. The pseudo-observations are produced by GC-Adjoint-HEMCO forward simulation. The full modeled CO fields are used in OSSE-FullOBS as pseudo-CO observations. The modeled CO fields are smoothed with MOPITT averaging kernels to produce MOPITT-like pseudo-CO observations in OSSE-MOPITT.

		<b>Asia</b>	<b>North America</b>	<b>Africa</b>	<b>South America</b>	<b>Europe</b>	<b>Australia</b>
<b>This work</b>	<b>A priori CO emissions</b>	<b>243.53</b>	<b>34.42</b>	<b>23.24</b>	<b>30.39</b>	<b>25.94</b>	<b>2.02</b>
	<b>A posteriori CO emissions</b>	<b>283.20</b>	<b>50.47</b>	<b>35.34</b>	<b>42.92</b>	<b>41.62</b>	<b>2.79</b>
	<b>Scaling Factors</b>	<b>1.16</b>	<b>1.47</b>	<b>1.52</b>	<b>1.41</b>	<b>1.60</b>	<b>1.38</b>
<b>Jiang et al. 2017</b>	<b>A priori CO emissions</b>	<b>270.50</b>	<b>43.70</b>	<b>29.39</b>	<b>17.47</b>	<b>44.45</b>	<b>0.83</b>
	<b>A posteriori CO emissions</b>	<b>205.40</b>	<b>47.06</b>	<b>35.04</b>	<b>16.67</b>	<b>53.58</b>	<b>0.82</b>
	<b>Scaling Factors</b>	<b>0.76</b>	<b>1.08</b>	<b>1.19</b>	<b>0.95</b>	<b>1.21</b>	<b>0.99</b>

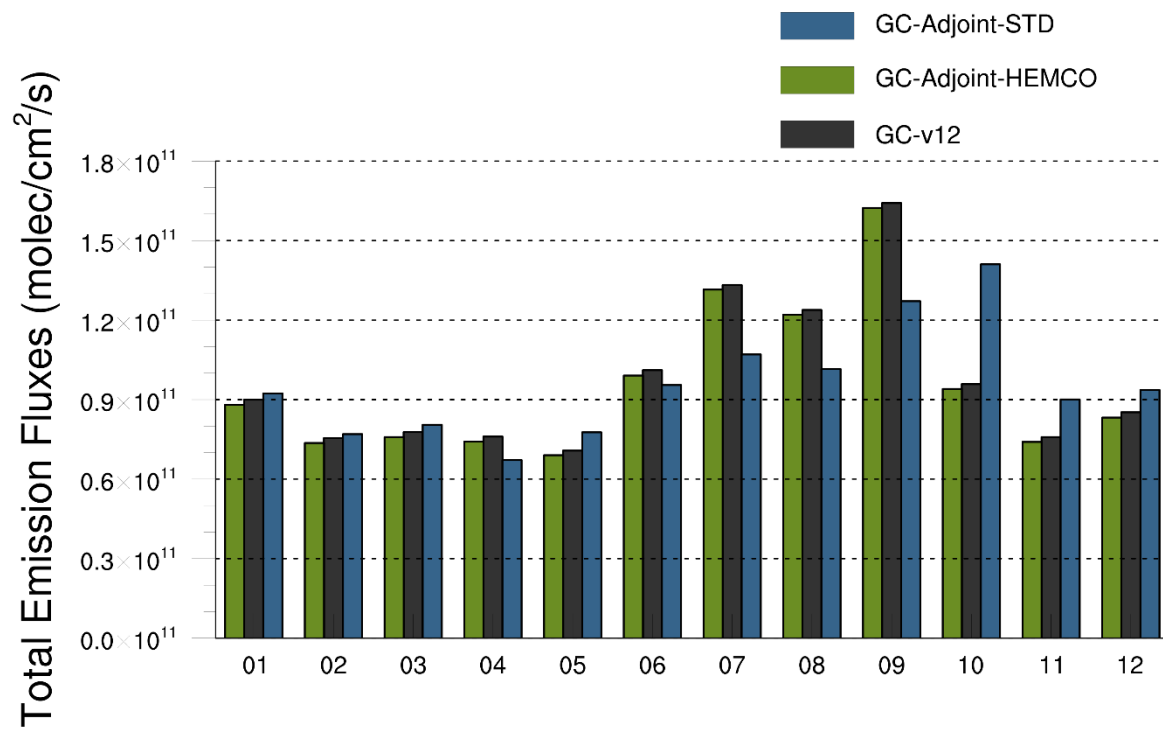
**Table 4.** Regional anthropogenic CO emissions (with unit Tg/y) and annual scaling factors in 2015 in this work and Jiang et al. 2017.



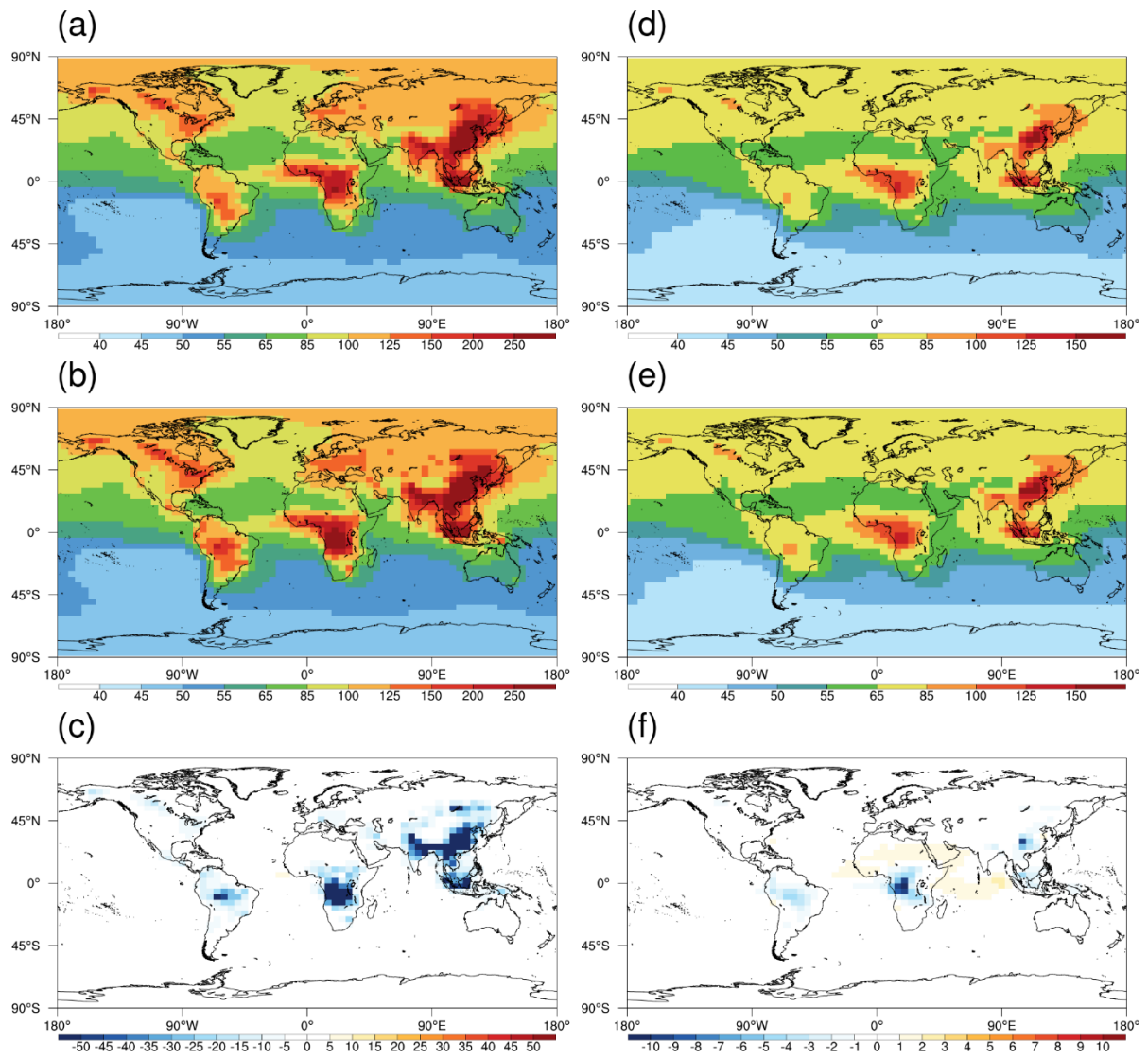
**Fig. 1.** Framework to read the updated emission inventories in GC-Adjoint-HEMCO.



**Fig. 2.** Total combustion CO emissions in 2015 from (a) GC-v12; (b) GC-Adjoint-HEMCO; (c) GC-Adjoint-STD. The unit is  $\text{molec}/\text{cm}^2/\text{s}$ .

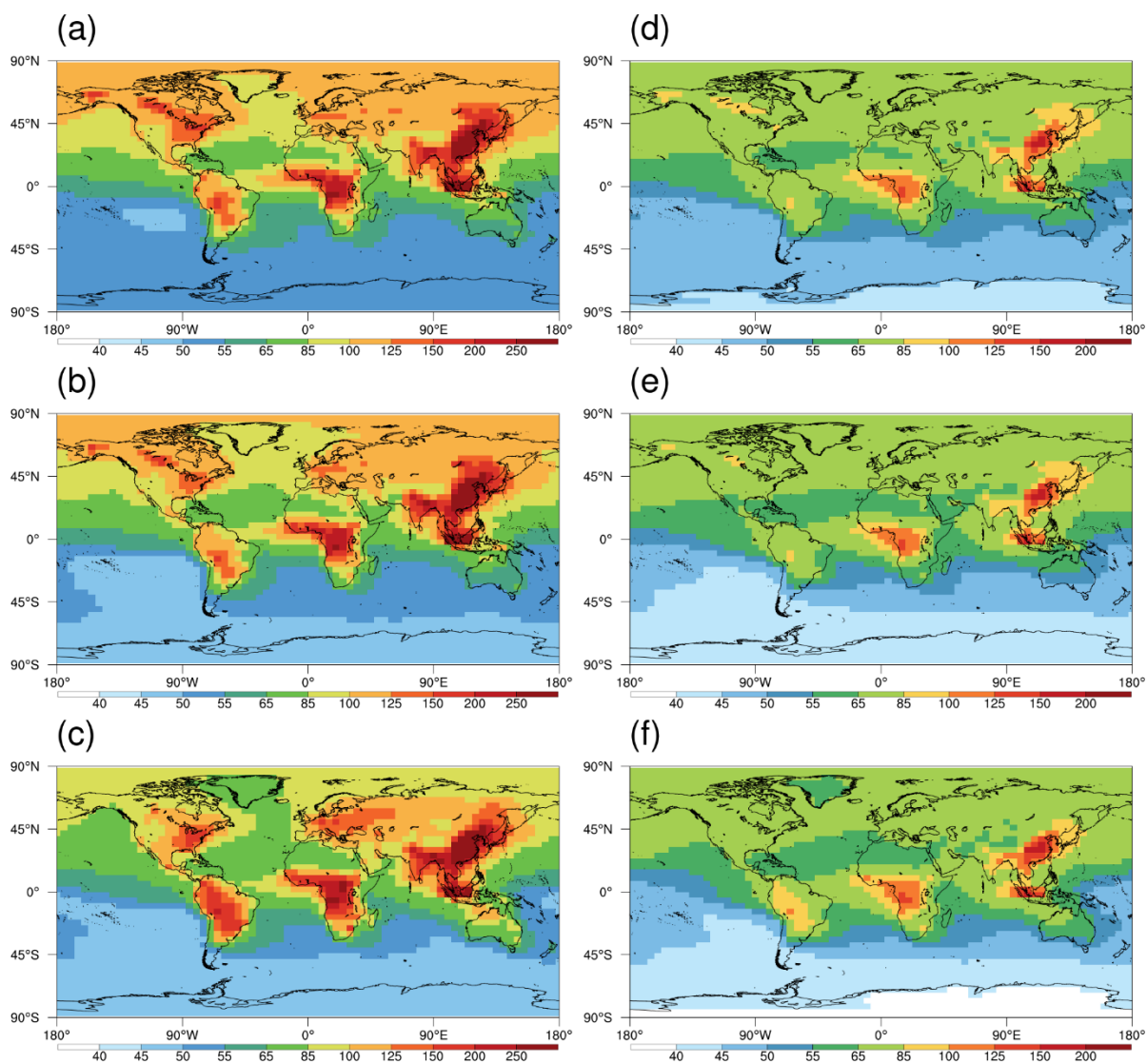


**Fig. 3.** Monthly variation in combustion CO emissions in 2015 from GC-v12, GC-Adjoint-HEMCO and GC-Adjoint-STD.

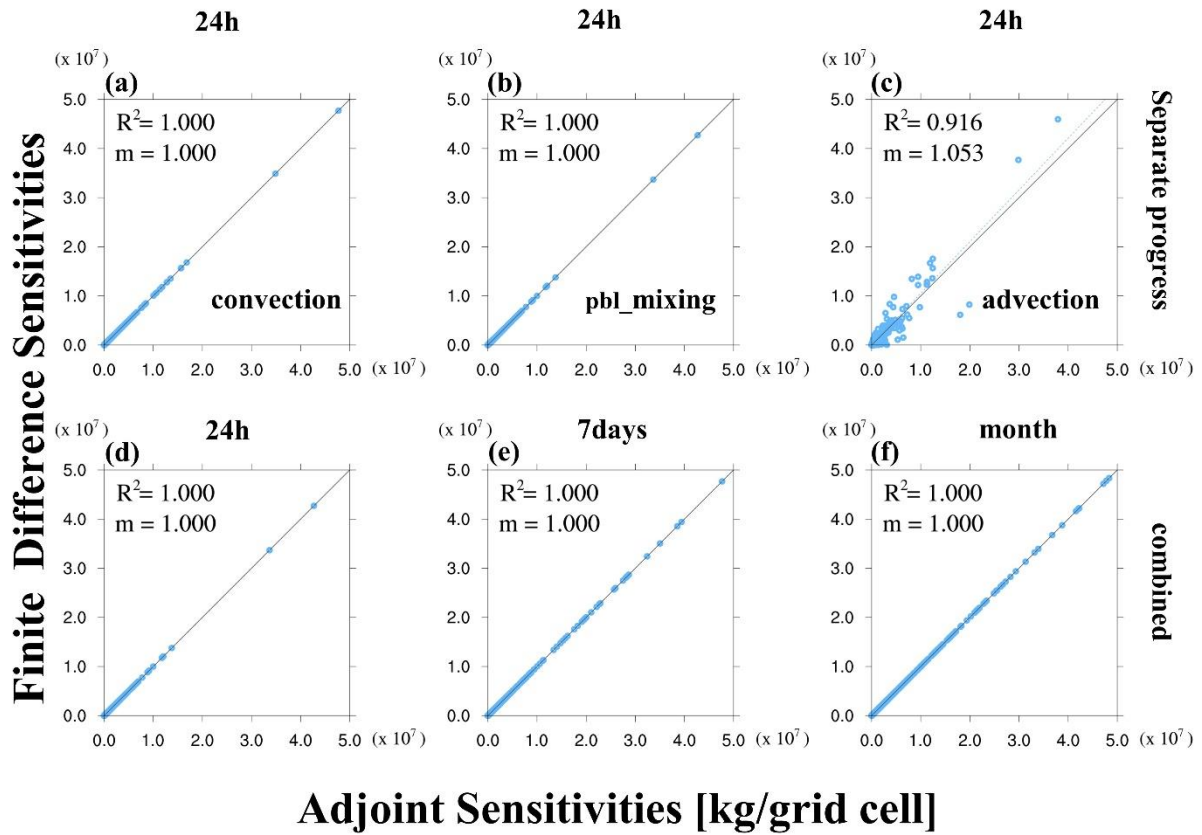


**Fig. 4.** Averages of surface CO concentrations (unit ppbv) in 2015 from (a) GC-Adjoint-HEMCO driven by MERRA-2, (b) GC-Adjoint-HEMCO driven by GEOS-FP and (c) their difference; (d-f) same as panels a-c, but for CO columns (column-averaged dry-air mole fractions,  $X_{CO}$ ).

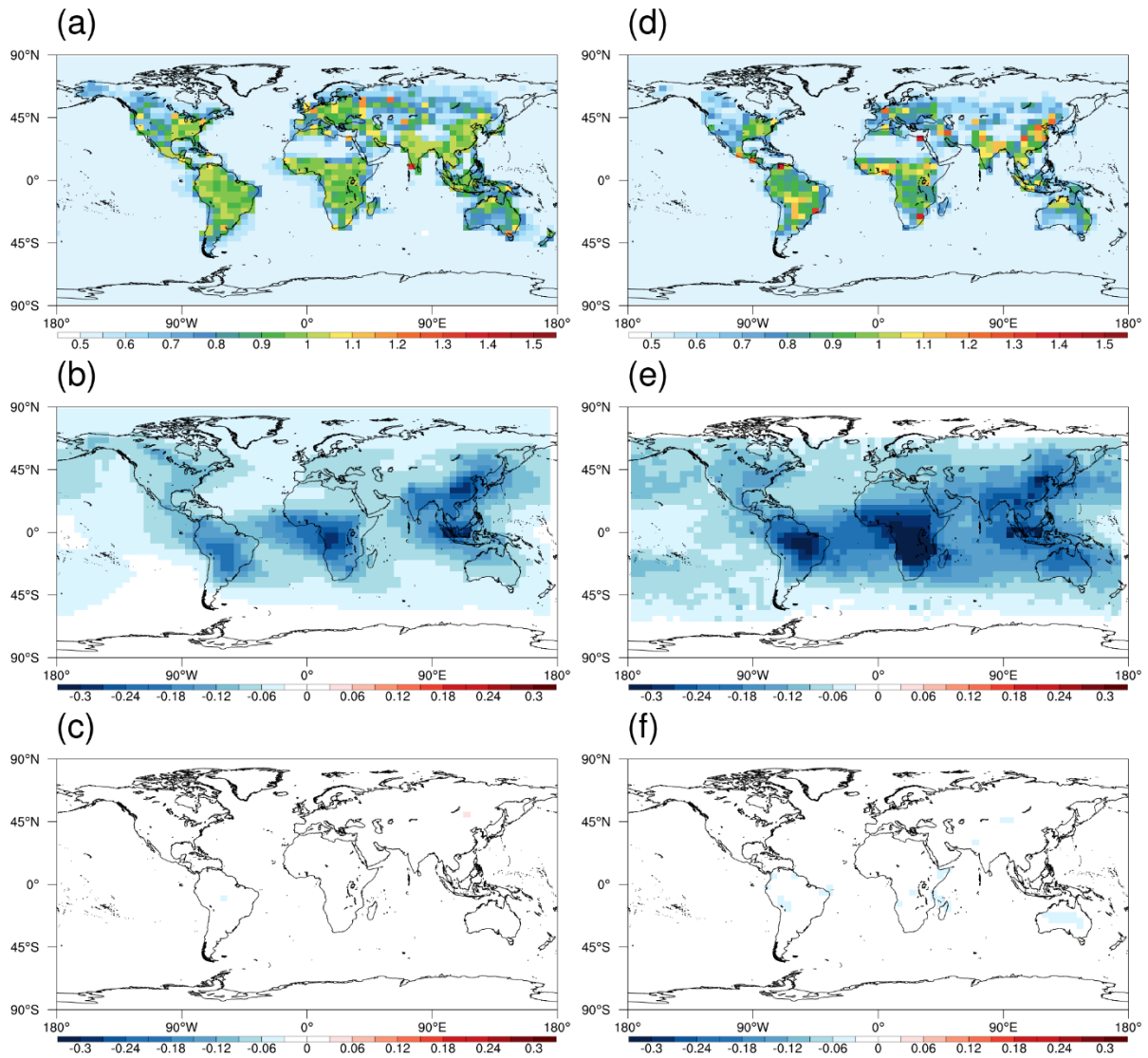




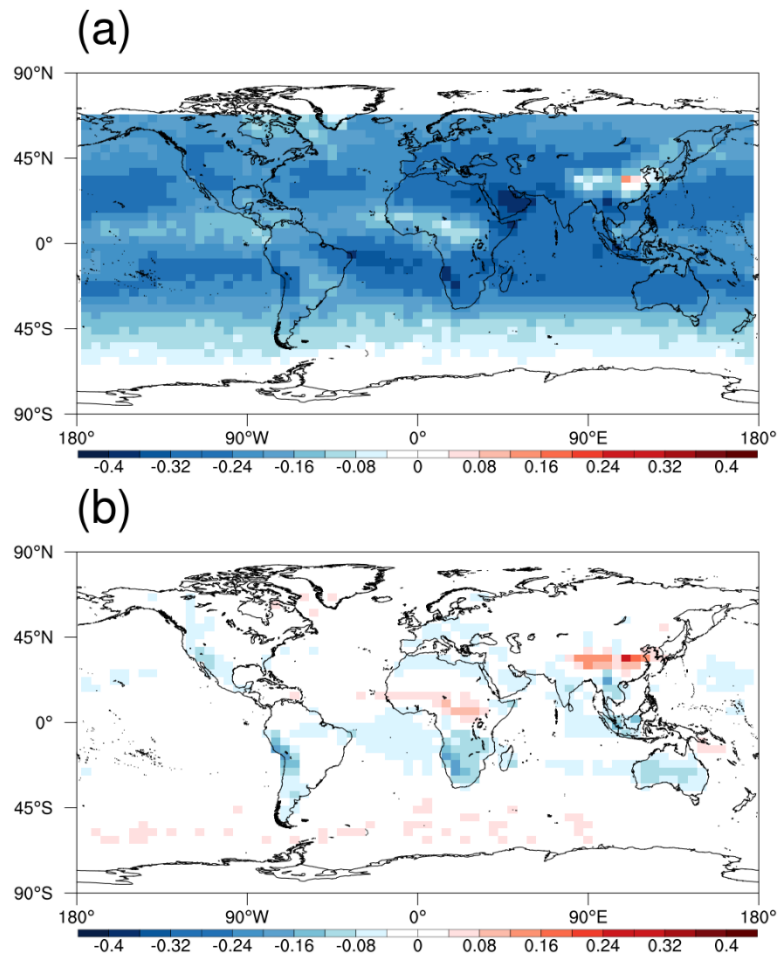
**Fig. 5.** Averages of surface CO concentrations (unit ppbv) in 2015 from (a) GC-v12; (b) GC-Adjoint-HEMCO; (c) GC-Adjoint-STD; (d-f) same as panels a-c, but for CO columns (column-averaged dry-air mole fractions,  $X_{co}$ ).



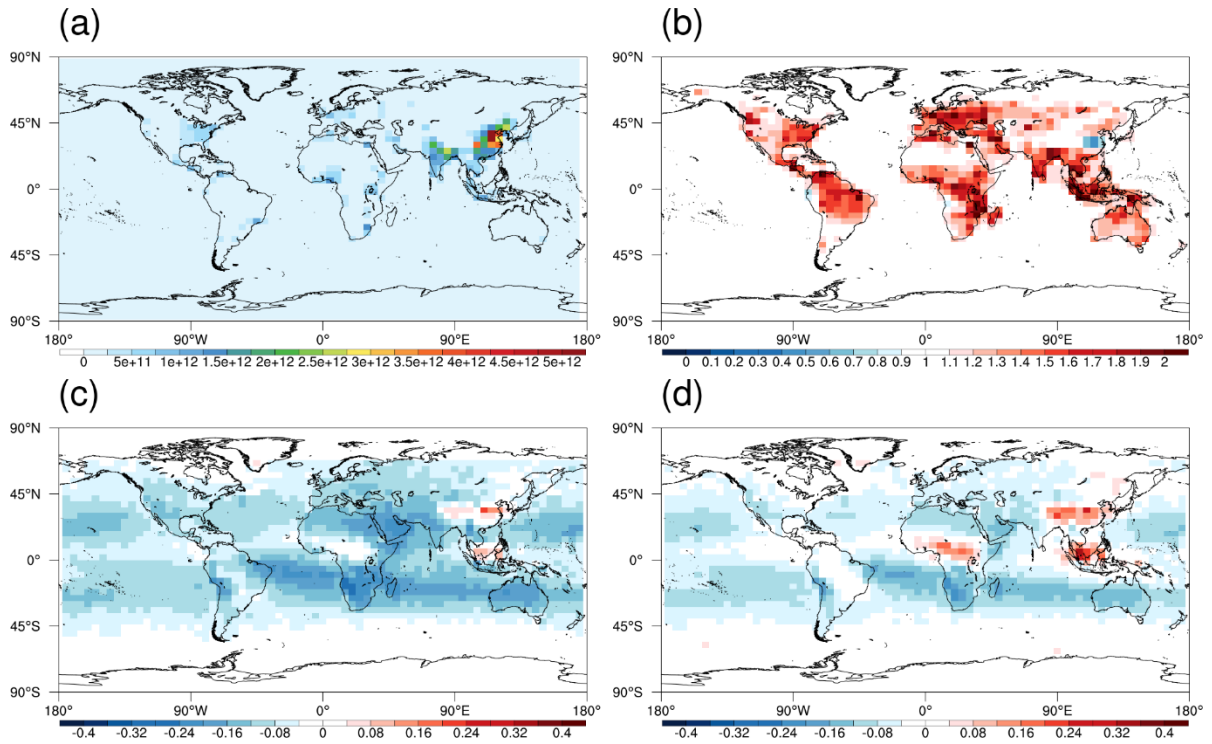
**Fig. 6.** Comparison of sensitivities of global CO concentrations (LFD\_GLOB and model level 1) to CO emission scaling factors calculated using the adjoint method vs. the finite difference method. (a-c) the effects of convection, PBL mixing and advection with 24-hour assimilation window; (d-f) the combined effects (the advection process is turned off) with increased assimilation windows.



**Fig. 7.** (a) Annual scaling factors in OSSE-FullOBS. The scaling factors represent the ratio of the estimated to true emissions. The ratio for the first guess is 0.5. The actual value is 1.0. (b-c) the a priori and a posteriori biases calculated by (model-observation)/observation in OSSE-Full. (d-f) same as panels a-c, but for OSSE-MOPITT.



**Fig. 8.** (a) Biases in monthly initial CO conditions in 2015 in the original GEOS-Chem simulation. (b) same as panel a, but with optimized initial CO conditions provided by suboptimal sequential Kalman filter. The biases are calculated by  $(\text{model-MOPITT})/\text{MOPITT}$ .



**Fig. 9.** (a) A priori anthropogenic CO emissions in 2015 with unit molec/cm<sup>2</sup>/s; (b) Annual scaling factors for CO emissions in 2015. The scaling factors represent the ratio of the estimated to true emissions. (c-d) the a priori and a posteriori biases calculated by (model-MOPITT)/MOPITT.