

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

3
4 Zhaojun Tang¹, Zhe Jiang^{1*}, Jiaqi Chen¹, Panpan Yang¹, Yanan Shen¹

5
6 ¹School 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)

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.

Deleted: Adjoint

Deleted: by supporting

Deleted: Zhe Jiang^{1*}

Deleted: various

Deleted: pollutants

Deleted: provide an updated version (GC-Adjoint-HEMCO) of ...

Deleted: the MERRA-2 meteorological data and

Deleted: . State-of-

Deleted: -art inventories, such as CEDS (Community Emissions Data System), MIX, NEI2011 (National Emissions Inventory), and GFED4 (Global Fire Emission Database), are supported in GC-Adjoint-HEMCO. We find good agreements in ...

Deleted: emissions from various inventories, chemical

Deleted: , and

Deleted: between GC-Adjoint-HEMCO

Deleted: GEOS-Chem (v12-8-1) forward simulations. Furthermore, ...

Deleted: are employed

Deleted: of GC-Adjoint-HEMCO

Deleted: We find underestimations by approximately 15% in the a posteriori anthropogenic CO emissions over North America and Europe due to limited coverage of observations by smoothing the pseudo-CO observations with Measurement of Pollution in the Troposphere (MOPITT) averaging kernels. As...

Deleted: GC-Adjoint-HEMCO, we

Deleted: The a posteriori anthropogenic CO emission estimates derived in this work match well with in North America and Africa but are overestimated in Asia, South America and Australia, which could be associated with the different treatment of MOPITT CO observations over ocean grids and the large differences in CO chemical sources and sinks. The updated model

Deleted: is a useful extension

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

70 1. Introduction

71 GEOS-Chem is a global 3D chemical transport model (CTM) and has been widely used
72 to analyze the sources and variabilities of atmospheric compositions (Whaley et al., 2015; Li
73 et al., 2019; Hammer et al., 2020; Jiang et al., 2022). GEOS-Chem model is driven by
74 meteorological reanalysis data from the Goddard Earth Observing System (GEOS) of the
75 Global Modeling and Assimilation Office (GMAO). Emissions in GEOS-Chem model are
76 calculated with state-of-the-art inventories such as CEDS (Community Emissions Data
77 System) (Hoesly et al., 2018), MIX (Li et al., 2017) and NEI2011 (National Emissions
78 Inventory). Based on GEOS-Chem forward simulation, the adjoint of the GEOS-Chem model
79 (Henze et al., 2007) further provides the capability of backward simulation of physical and
80 chemical processes within the 4D variational (4D-var) framework. The major advantage of the
81 adjoint model is obtaining the sensitivity of atmospheric concentrations to multiple model
82 variables within a single backward simulation. The major applications of the adjoint of GEOS-
83 Chem model include inverse analyses of atmospheric composition emissions by minimizing
84 the difference between simulations and observations (Jiang et al., 2015a; Zhang et al., 2018;
85 Qu et al., 2022) as well as sensitivity analyses to analyze the sources of atmospheric
86 compositions (Jiang et al., 2015b; Zhao et al., 2019; Dedoussi et al., 2020).

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

Deleted: various

Moved down [1]: Harmonized Emissions Component (HEMCO) (Keller et al., 2014; Lin et al., 2021)

Deleted: based on the

Deleted: with default support for

Deleted: model

Deleted: pollutant

Deleted: pollutants

Deleted: backward simulation

Deleted: code

Deleted: backward code

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

113 Emission inventories play a key role in the simulation of atmospheric compositions,
114 Harmonized Emissions Component (HEMCO) (Keller et al., 2014; Lin et al., 2021) was
115 included in the GEOS-Chem forward simulations in v10-01. HEMCO is responsible for inputs
116 of meteorological and emission data with default support for emission inventories such as
117 CEDS, MIX and NEI2011. New emission inventories can be added readily within HEMCO
118 framework. There are noticeable differences between HEMCO and the adjoint of GEOS-Chem
119 model. First, meteorological and emission data are read with individual modules in the adjoint
120 of GEOS-Chem model. Second, the inputs of emission inventories are undertaken by different
121 modules that were developed individually with significant discrepancies in the source code. In
122 addition, the file format (e.g., binary punch in the adjoint of GEOS-Chem that is the format of
123 older GEOS-Chem versions in contrast to netCDF in HEMCO), emission variables and the
124 usage methods of emission variables (e.g., emission hierarchy, scaling factors and time slice)
125 are inconsistent. These differences have posed a barrier to the application of new emission
126 inventories in the adjoint of GEOS-Chem model.

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

Deleted:), which can lead to discontinuity in the derived trends of emissions . Furthermore, the HEMCO emission module was included in the GEOS-Chem forward simulations in v10-01. Adjoint of GEOS-Chem model does not support the HEMCO module, and thus, updated emission inventories such as CEDS, MIX and NEI2011 cannot be read conveniently, which can affect the performance in the simulated atmospheric compositions

Moved (insertion) [1]: Harmonized Emissions Component (HEMCO) (Keller et al., 2014; Lin et al., 2021)

Field Code Changed

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

149 Ideally, people should consider porting the complete HEMCO to the adjoint of GEOS-
150 Chem model to match the new features in GEOS-Chem forward simulations. However, a
151 complete port of HEMCO implies replacing the input framework of the adjoint of GEOS-Chem
152 model, as well as restructuring of HEMCO and the adjoint of GEOS-Chem model to address
153 the compatibility issues, which is very challenging and may not be necessary because the
154 meteorological modules still work well in the adjoint of GEOS-Chem model. Consequently, a
155 major objective of this work is to design a new framework to facilitate emission inventory
156 updates in the adjoint of GEOS-Chem model. For this objective, this new framework must have
157 good readability and extensibility to allow us to support HEMCO emission inventories
158 conveniently and to add more emissions inventories following future updates in GEOS-Chem
159 forward simulations easily. Furthermore, we developed new modules to support MERRA-2
160 meteorological data within the current framework of the adjoint of GEOS-Chem model, as
161 reuse of existing frameworks can save much work.

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

Deleted: In this work, we develop the capability of the adjoint of GEOS-Chem model to support MERRA-2 meteorological data and HEMCO emission inventories. The results presented in this paper show the development, evaluation, and application of the developed capability to constrain carbon monoxide (CO) emissions by assimilating CO measurements from the Measurement of Pollution in the Troposphere (MOPITT).

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

182 The results presented in this paper show the development, integration, evaluation, and
183 application of these new capabilities, which is important to better applications of the adjoint of
184 GEOS-Chem model in the future. The capabilities developed in this work will be submitted to
185 the standard GEOS-Chem adjoint code base (Henze et al., 2007) for better development of the
186 community of the adjoint of GEOS-Chem model. This paper is organized as follows: in Section
187 2, we describe the adjoint of GEOS-Chem model, the development of these new capabilities,
188 and the Measurement of Pollution in the Troposphere (MOPITT) CO observations used in this
189 work. In Section 3, we evaluated the performances of the developed capabilities in forward and
190 backward simulations, together with observing system simulation experiments (OSSE) to
191 evaluate the model performance in 4D-var assimilations. An example application of 4D-var
192 assimilation to constrain anthropogenic CO emissions in 2015 by assimilating MOPITT CO
193 observations, was also presented. Our conclusions follow in Section 4.

195 2. Methodology and Data

196 2.1 Adjoint of the GEOS-Chem model

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

Deleted: capability presented

Deleted: is

Deleted: the

Deleted: capability to support the MERRA-2 meteorological data...

Deleted: HEMCO emission inventories, and the

Deleted: performance

Deleted: and presented an

Deleted: .

207 a horizontal resolution of $4^\circ \times 5^\circ$ with 47 vertical levels and employs the CO-only simulation
 208 (tagged-CO mode). The global default anthropogenic emission inventory in the standard
 209 version of the adjoint of GEOS-Chem model (hereafter referred to as GC-Adjoint-STD) is
 210 Global Emissions Initiative (GEIA), but is replaced by the following regional emission
 211 inventories: NEI2008 in North America, the Criteria Air Contaminants (CAC) inventory for
 212 Canada, the Big Bend Regional Aerosol and Visibility Observational (BRAVO) Study
 213 Emissions Inventory for Mexico (Kuhns et al., 2003), the Cooperative Program for Monitoring
 214 and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP) inventory
 215 for Europe in 2000 (Vestreng and Klein, 2002) and the INTEX-B Asia emissions inventory for
 216 2006 (Zhang et al., 2009). Biomass burning emissions are based on the GFED3 (van der Werf
 217 et al., 2010).

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

$$220 \quad J(\mathbf{x}) = \sum_{i=1}^N (\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i)^T \mathbf{S}_\Sigma^{-1} (\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i) + \gamma (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) \quad (1)$$

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

$$227 \quad \nabla_{\mathbf{x}} J(\mathbf{x}) = \sum_{i=1}^N \left[2\mathbf{S}_\Sigma^{-1} (\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i) \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}} \right] + 2\gamma \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) \quad (2)$$

228 We assume a uniform observation error of 20%. The combustion CO sources (fossil fuel,
 229 biofuel and biomass burning) and the oxidation source from biogenic volatile organic
 230 compounds (VOCs) are combined, assuming a 50% uniform a priori error. We optimize the
 231 source of CO from the oxidation of methane (CH₄) separately as an aggregated global source,

Deleted: by adjusting the CO emissions iteratively

Deleted: $J(\mathbf{x}) = \sum_{i=1}^N [\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i]^T \mathbf{S}_\Sigma^{-1} [\mathbf{F}_i(\mathbf{x}) - \mathbf{z}_i] + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a)$ (1)

Deleted: \mathbf{x}

Deleted: \mathbf{z}_i

Deleted: $\mathbf{F}(\mathbf{x})$

Deleted: \mathbf{S}_Σ

Deleted: \mathbf{S}_a

Deleted: compound (VOC)

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

Deleted: We

Deleted: and

248 2.2 ~~New framework to read~~ emission inventories

Deleted: Updates in

249 A major objective of this work is to ~~design a new framework to facilitate emission~~
250 ~~inventory updates in~~ the adjoint of GEOS-Chem model. As shown in Fig. 1, we first initialize
251 the array in [INITIAL] and batch read the emission data in [READ_DATA], which were
252 interpolated offline with $1^\circ \times 1^\circ$ resolution by considering the mass conservation. Here, the data
253 include the emission inventory data listed in Table S1 (see the SI), the corresponding scaling
254 factor data and the mask map files of domain definitions. The data are scaled in
255 [SCALE_DATA] by multiplying the corresponding annual, season, month, week, and 24-hour
256 emission factors and are then online interpolated to the current resolution ($4^\circ \times 5^\circ$ in this work)
257 of the model by [RGRID_DATA], which was followed by the application of region masks in
258 [MASK].

Deleted: obtain consistent emissions between

Deleted: and the GEOS-Chem forward model (v12-08-01, hereafter referred to as GC-v12).

259 The emission variable of CO obtained in this part is written to the model memory in
260 emission.f and emission_adj.f by calling DO_EMISSIONS ~~to ensure the consistent emissions~~
261 ~~in both forward and backward simulations~~. The GET_[TRACER] subroutines are used to
262 obtain the CO emission variable, which participates in the calculation of physicochemical
263 processes in the model, to interact with other modules. Finally, the variable is cleaned from the
264 memory by the [CLEANUP] module. It should be noted that a two-step interpolation is
265 employed in ~~this work (hereafter referred to as GC-Adjoint-HEMCO)~~ following GC-Adjoint-

Deleted: .

273 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
274 different from the one-step interpolation in GEOS-Chem forward model (v12-08-01, hereafter
275 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.
276 The different interpolation methods can lead to differences in the interpolated emission data.

Deleted: GC-v12,

277 2.3 Updates in emission inventories

278 In addition to baseline emission data, there are critical factors that affect the usage of
279 emission data in the models. Reading the emission data correctly thus does not necessarily
280 mean using emission data correctly. For example, emission hierarchy is used to prioritize
281 emission fields within the same emission category. Emissions of higher hierarchy overwrite
282 lower hierarchy data. Regional emission inventories usually have a higher hierarchy within
283 their mask boundaries. Scaling factors are used to adjust the baseline emissions with annual,
284 season, month, week, and 24-hour temporal scales. Time slice selection is used to define the
285 usage methods of the emission data outside the original temporal range; for instance, data can
286 be interpreted as climatology and recycled once the end of the last time slice is reached or be
287 only considered as long as the simulation time is within the time range. Furthermore, there are
288 experience parameters applied in files such as emfossil.f and tagged_co.f, which may not be
289 compatible with HEMCO emission inventories. Consequently, we must validate the integrated
290 emissions carefully to ensure that the abovementioned factors have been correctly applied and
291 to ensure that the calculated emissions are reasonable for individual inventories and the
292 combination of all inventories.

293 To take advantage of this new framework, six HEMCO emission inventories have been
294 added to this work. To validate the emissions, we performed actual simulations with GC-v12,
295 GC-Adjoint-HEMCO and GC-Adjoint-STD, and the emissions were calculated in the model
296 simulations and then output to the Log file. As shown in Table S1, the CEDS emission
297 inventory ($0.5^\circ \times 0.5^\circ$) is adopted in GC-Adjoint-HEMCO to provide global default emissions

299 for 1750-2019. The diurnal scale factors are applied to obtain CO emissions at different
300 moments of the day. Fig. S1 (see the SI) shows CEDS CO emissions in 2015 in GC-v12 and
301 GC-Adjoint-HEMCO and GEIA CO emissions in GC-Adjoint-STD, and we find noticeable
302 differences in CO emissions between CEDS and GEIA. As shown in Table 1, the CEDS CO
303 emissions in 2015 were 613.57 and 613.85 Tg/y in GC-v12 and GC-Adjoint-HEMCO,
304 respectively, with a relative difference of 0.05% between GC-v12 and GC-Adjoint-HEMCO.
305 The GEIA CO emissions in 2015 were 445.88 Tg/year in GC-Adjoint-STD.

306 The default CEDS inventory is replaced by the following regional emission inventories
307 in GC-Adjoint-HEMCO: MIX in Asia ($0.25^{\circ} \times 0.25^{\circ}$), NEI2011 in the United States
308 ($0.1^{\circ} \times 0.1^{\circ}$), DICE_AFRICA and EDGARV43 in Africa ($0.1^{\circ} \times 0.1^{\circ}$) and APEI in Canada
309 ($0.1^{\circ} \times 0.1^{\circ}$). As shown in Fig. S2 (see the SI), the MIX inventory provides Asian emissions in
310 2008-2010, accompanied by diurnal scale factors to describe daily emission variation. The
311 $1^{\circ} \times 1^{\circ}$ scale factors in the AnuualScalar.geos.1x1.nc file further provide the annual variation in
312 1985-2010. As shown in Table 1, the MIX CO emissions in 2015 were 321.18 and 321.71 Tg/y
313 in GC-v12 and GC-Adjoint-HEMCO, respectively, with a relative difference of 0.17% between
314 GC-v12 and GC-Adjoint-HEMCO. The INTEX-B CO emissions in 2015 were 353.03 Tg/y in
315 GC-Adjoint-STD.

316 The NEI2011 inventory (Fig. S3, see the SI) provides anthropogenic emissions for the
317 United States in 2011 with annual scalar factors from 2006-2013. The weekday and weekend
318 factors are read from NEI99.dow.geos.1x1.nc file since 1999 with all CO factors of 1.0 on
319 weekdays and between 0.990 and 0.997 on Saturdays and Sundays. The NEI2011 CO
320 emissions in 2015 were 35.83 and 37.70 Tg/y in GC-v12 and GC-Adjoint-HEMCO,
321 respectively, with a relative difference of 5.22% between GC-v12 and GC-Adjoint-HEMCO.
322 The NEI2008 CO emissions in 2015 were 52.87 Tg/y in GC-Adjoint-STD. APEI (Fig. S4, see
323 the SI) is the primary source of anthropogenic emissions in the Canadian domain. The APEI

Deleted: mid-week

325 CO emissions in 2015 were 6.10 and 6.17 Tg/y in GC-v12 and GC-Adjoint-HEMCO,
326 respectively, with a relative difference of 1.14% between GC-v12 and GC-Adjoint-HEMCO.
327 The CAC CO emissions in 2015 were 10.20 Tg/y in GC-Adjoint-STD. Following GC-v12, the
328 CO emissions in APEI are enhanced by 19% to account for coemitted VOC in the tagged-CO
329 simulation.

330 Emissions for the African domain are provided by the combination of DICE_AFRICA
331 and EDGARV43 (Fig. S5, see the SI). Here DICE_AFRICA includes anthropogenic and
332 biofuel emissions in 2013. We read the DICE_AFRICA emissions data into the model in two
333 types according to the guidelines of the inventory. Emissions from sectors such as automobiles
334 and motorcycles are aggregated into anthropogenic sources, and household-generated
335 emissions such as charcoal and agricultural waste are aggregated into biofuel sources. Efficient
336 combustion emissions from EDGAR v4.3 in 1970-2010 then compensate for the lacking
337 sources in DICE_AFRICA. Daily variation factors for CO are also used here for emissions
338 across the African region. The 2010 CO seasonal scale factors are used in EDGAR v4.3 for
339 sectoral emission sources. The DICE_AFRICA and EDGARV43 CO emissions in 2015 were
340 83.42 and 83.02 Tg/y in GC-v12 and GC-Adjoint-HEMCO, respectively, with a relative
341 difference of -0.48% between GC-v12 and GC-Adjoint-HEMCO. Following GC-v12, the CO
342 emissions in DICE_AFRICA and EDGARV43 are enhanced by 19% to account for coemitted
343 VOC in the tagged-CO simulation.

344 The biomass burning emission inventory in GC-Adjoint-HEMCO is GFED4 (Fig. S6,
345 see the SI), which includes dry matter emissions from a total of seven sectors in 1997-2019.
346 The same GFED_emssion_factors.H header file as in the GC-v12 version is read in the GC-
347 Adjoint-HEMCO. This file contains the ratio factors of atmospheric pollutants, and we
348 multiply the ratio factors one by one according to the ID of each species to ensure that the
349 species in the model have biomass burning sources. The GFED4 CO emissions in 2015 were

Deleted: the

351 437.13 and 435.89 Tg/y in GC-v12 and GC-Adjoint-HEMCO, respectively, with a relative
352 difference of -0.28% between GC-v12 and GC-Adjoint-HEMCO. The GFED3 CO emissions
353 in 2015 were 382.04 Tg/year in GC-Adjoint-STD. Following GC-v12, the combustion CO
354 sources in biomass burning are enhanced by 5% to consider the CO generated by VOC in the
355 tagged-CO simulation.

356 Fig. 2 shows the total combustion CO emissions in 2015 from GC-v12, GC-Adjoint-
357 HEMCO and GC-Adjoint-STD. As shown in Table 2, the regional combustion CO emissions
358 are 320.66 and 320.38 Tg/y (Asia), 73.96 and 66.93 Tg/y (North America), 199.51 and
359 193.29 Tg/y (Africa), 79.04 and 78.91 Tg/y (South America), 31.58 and 30.96 Tg/y (Europe)
360 and 12.24 and 11.99 Tg/y (Australia) in GC-v12 and GC-Adjoint-HEMCO, respectively. Fig.
361 3 further shows the monthly combustion CO emissions in 2015 from GC-v12, GC-Adjoint-
362 HEMCO and GC-Adjoint-STD, and there are good agreements in the monthly variation of CO
363 emissions between GC-v12 and GC-Adjoint-HEMCO. The CO emissions in GC-Adjoint-STD
364 are similar to those in GC-v12 and GC-Adjoint-HEMCO in winter and spring but with large
365 differences in summer and autumn. This seasonal difference may reflect the influence of
366 different emission inventories on biomass burning.

367 **2.4 Updates in CO chemical sources and sinks**

368 The biogenic emissions in GC-Adjoint-STD are Model of Emissions of Gases and
369 Aerosols from Nature, version 2.0 (MEGANv2.0, Guenther et al. (2006)) in the full chemistry
370 simulation but are GEIA in the tagged-CO simulation (Fig. S7, see the SI). Fisher et al. (2017)
371 demonstrated improvement in modeled CO concentrations in tagged-CO simulation by reading
372 archived VOC- and CH₄-generated CO fields provided by full chemistry simulation. The
373 archived VOC- and CH₄-generated CO fields in 2013 (PCO_3Dglobal.geosp.4x5.nc) were set
374 as the default CO chemical sources in the tagged-CO simulation in GC-v12 and supported in
375 GC-Adjoint-HEMCO. As shown in Table 2, the CO chemical sources (columns) obtained by

Deleted: 3

377 reading the archived VOC- and CH₄-generated CO fields demonstrate good agreement between
378 GC-v12 and GC-Adjoint-HEMCO. However, they are 30-60% lower than those in GEIA in
379 GC-Adjoint-STD, and this difference could be partially associated with the inconsistency
380 between the archived VOC-generated CO fields in 2013 and the actual meteorological data in
381 2015 in the simulation.

382 The default CH₄-generated CO emissions in GC-Adjoint-STD (Fig. S8, see the SI) are
383 calculated based on averaged CH₄ concentrations in four latitude bands (90°S - 30°S, 30°S -
384 00°S, 00°N - 30°N, 30°N - 90°N), which are based on Climate Monitoring and Diagnostics
385 Laboratory (CMDL) surface observations and Intergovernmental Panel on Climate Change
386 (IPCC) future scenarios. As shown in Table 2, there are good agreements in the CH₄-generated
387 CO emissions between GC-v12 and GC-Adjoint-HEMCO by reading
388 PCO_3Dglobal.geosfp.4x5.nc, and they are 20-60% lower than those in CMDL/IPCC in GC-
389 Adjoint-STD. Furthermore, the default archived monthly OH fields were updated following
390 GC-v12 with updated calculations for the decay rate (KRATE, from JPL 03 to JPL 2006) in
391 GC-Adjoint-HEMCO. The subsequent CO sinks (Fig. S9, see the SI) in GC-v12 and GC-
392 Adjoint-HEMCO are 20-40% higher than those in GC-Adjoint-STD.

393 **2.5 Updates in meteorological data**

394 The MERRA-2 meteorological data (1979-present) are supported in GC-Adjoint-
395 HEMCO to ensure long-term consistency in the meteorological data in the analyses. The code
396 porting to support MERRA-2 follows the current framework of the adjoint of GEOS-Chem
397 model, particularly because the meteorological variables and vertical resolutions of MERRA-
398 2 are the same as those of GEOS-FP (2012-present), while GEOS-FP is already supported by
399 GC-Adjoint-STD. Fig. 4A-B show the averages of surface CO concentrations in 2015 from
400 GC-Adjoint-HEMCO driven by MERRA-2 and GEOS-FP, respectively. Our results
401 demonstrate lower surface CO concentrations driven by MERRA-2 (Fig. 4C), although there

Deleted: 4

Deleted: inverse

Deleted: is more direct than emission inventories

405 is good agreement in the spatial distributions of CO concentrations. Similarly, Fig. 4D-F show
406 the averages of CO columns in 2015 from GC-Adjoint-HEMCO driven by MERRA-2 and
407 GEOS-FP and their differences. Despite the noticeable differences in surface CO
408 concentrations (Fig. 4C), the differences in CO columns (Fig. 4F) are much smaller, and the
409 modeled CO columns driven by MERRA-2 are higher than those driven by GEOS-FP over the
410 Indian Ocean. The discrepancy between surface and column CO in Fig. 4 may reflect the
411 impacts of different convective transports on the modeled CO concentrations.

412 **2.6 MOPITT CO measurements**

413 The MOPITT data used here were obtained from the joint retrieval (V7J) of CO from
414 thermal infrared (TIR, 4.7 μ m) and near-infrared (NIR, 2.3 μ m) radiances using an optimal
415 estimation approach (Worden et al., 2010; Deeter et al., 2017). The retrieved volume mixing
416 ratios (VMR) are reported as layer averages of 10 pressure levels with a footprint of 22 km \times
417 22 km. Following Jiang et al. (2017), we reject MOPITT data with CO column amounts less
418 than 5×10^{17} molec/cm² and with low cloud observations. Since the NIR channel measures
419 reflected solar radiation, only daytime data are considered.

420

421 **3. Model evaluation and application**

422 **3.1 Model performances in forward and backward simulations**

423 The reasonable emissions in the diagnostic outputs in Section 2 do not necessarily mean
424 the correct integration of emissions in the assimilations. Consequently, here we evaluate the
425 performance of GC-Adjoint-HEMCO in forward simulations. Fig. 5 shows the averages of
426 surface and column CO concentrations in 2015 from GC-v12, GC-Adjoint-HEMCO and GC-
427 Adjoint-STD. As shown in Table 2, the regional differences between GC-v12 and GC-Adjoint-
428 HEMCO are 2.6%, -5.7%, -4.6%, -1.7%, -1.4% and -3.6% in surface CO concentrations, and
429 -2.3%, -3.6%, -3.3%, -3.1%, -3.3% and -4.1% in CO columns over Asia, North America,

Deleted: 5

Deleted: CO concentrations

Deleted: the

Deleted: We first

434 Africa, South America, Europe, and Australia, respectively. There are larger regional
435 differences in CO concentrations between GC-v12 and GC-Adjoint-STD; 4.6%, -10.1%, 6.3%,
436 22.5%, 6.4% and 25.7% in surface CO concentrations, and -0.7%, -9.9%, 2.5%, 8.0%, -5.8%
437 and 8.5% in CO columns over Asia, North America, Africa, South America, Europe, and
438 Australia, respectively. The agreement between GC-v12 and GC-Adjoint-HEMCO confirms
439 the reliability of GC-Adjoint-HEMCO in forward simulations, while the small differences in
440 CO concentrations between GC-v12 and GC-Adjoint-HEMCO are expected in view of the
441 comparable differences in regional emissions, chemical sources and sinks, as shown in Table
442 2.

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

$$449 \quad -\nabla_x J(x) = \frac{\partial F_N}{\partial x} \quad (3)$$

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

$$453 \quad \Lambda = \frac{J(\sigma+\delta\sigma)-J(\sigma-\delta\sigma)}{2\delta\sigma} \quad (4)$$

454 Here the finite difference gradients represent the response of modeled atmospheric
455 compositions at the final time step to finite perturbations in emissions provided by the forward
456 simulations ($\delta\sigma = 10\%$ in this work).

457 Fig. 6A-C show the comparison of adjoint and finite difference gradients of global CO
458 concentrations to CO emissions with a 24-hour assimilation window by turning on the

Deleted: The differences in CO concentrations between GC-v12 and GC-Adjoint-HEMCO are expected in view of the comparable differences in regional emissions, chemical sources and sinks, as shown in Table 2. The

Deleted: are larger

Moved (insertion) [2]: Fig.

465 convection, planetary boundary layer mixing and advection processes individually. We find
466 good consistency in the gradients with respect to convection and planetary boundary layer
467 mixing. The larger deviation with respect to advection is caused by the discrete advection
468 algorithm in forward simulations and continuous advection algorithm in backward simulations
469 (Henze et al., 2007). Fig. 6D-F further exhibit the effects of combined model processes (turning
470 off advection as suggested by Henze et al. (2007)). We find good agreement between the
471 adjoint and finite difference gradients with different assimilation windows (24 hours, 7 days
472 and one month). This confirms the consistency in the impacts of emissions to modeled
473 atmospheric compositions between the forward and backward simulations, which is the
474 prerequisite for more detailed evaluations in the following Sections.

475 **3.2 Observing system simulation experiments with pseudo-CO observations**

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

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

Deleted: as a 4D-var framework using observing system simulation experiments (OSSE).

492 and time step (hereafter referred to as OSSE-FullOBS). This experiment is designed to evaluate
493 the performance of the assimilation system under ideal conditions with full coverage of
494 observations. 2) The modeled CO fields are sampled at the locations/times of MOPITT CO
495 observations and smoothed with MOPITT a priori concentrations and averaging kernels to
496 produce MOPITT-like pseudo-CO observations (hereafter referred to as OSSE-MOPITT). This
497 experiment is designed to evaluate the performance of the assimilation system under actual
498 conditions with limited coverage of observations.

499 In the inverse analysis with the pseudo-CO observations, we reduce the anthropogenic
500 CO emissions by 50% so that the objective of the OSSE is to produce scaling factors that can
501 return the source estimate to the default emissions (i.e., scaling factors of 1.0). Fig. 7A shows
502 the annual scaling factors in 2015 in OSSE-FullOBS. After 40 iterations, the a posteriori
503 anthropogenic CO emission estimates converge to the true states in all major emission regions.
504 As shown in Table 3, the regional scaling factors of OSSE-FullOBS are 1.00, 0.97, 0.97, 1.00,
505 0.98 and 0.94 for anthropogenic CO emissions over Asia, North America, Africa, South
506 America, Europe, and Australia, respectively.

507 Furthermore, Fig. 7D shows the annual scaling factors in OSSE-MOPITT, which are
508 noticeably worse than those in Fig. 7A. The regional scaling factors of OSSE-MOPITT are
509 1.04, 0.88, 1.01, 1.02, 0.84 and 0.81 for anthropogenic CO emissions over Asia, North
510 America, Africa, South America, Europe, and Australia, respectively. With respect to OSSE-
511 FullOBS, the limited coverage of observations in OSSE-MOPITT has resulted in
512 approximately 15% underestimations in the a posteriori CO emission estimates over North
513 America and Europe. In addition, Fig. 7B-C and Fig. 7E-F show the a priori and a posteriori
514 biases in the modeled CO columns. We find dramatic improvements in the modeled CO
515 columns, which confirms the reliability of the 4D-var assimilation system. The difference
516 between Fig. 7B and 6E reflects the influence of the application of MOPITT averaging kernels,

Deleted: 6A

Deleted: 6D

Deleted: 6A

Deleted: 6B

Deleted: 6E

Deleted: 6B

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

526 3.3 Anthropogenic CO emissions constrained with MOPITT CO observations

527 As an example of the application of GC-Adjoint-HEMCO, here we constrain
528 anthropogenic CO emissions in 2015 by assimilating MOPITT CO observations. Fig. [8A](#) shows
529 the relative differences between modeled and MOPITT CO columns at the beginning of each
530 month in 2015 (i.e., biases in monthly initial CO conditions) in the original GEOS-Chem
531 simulations. We find dramatic underestimations in the modeled CO columns by approximately
532 30-40%. As indicated by previous studies (Jiang et al., 2013; Jiang et al., 2017), the biases in
533 monthly initial CO conditions are caused by model biases in CO concentrations accumulated
534 in previous months. Considering that the lifetime of CO is approximately 2-3 months, the
535 negative biases in the initial conditions can result in negative biases in the modeled CO
536 concentration in the following month. A lack of consideration of these biases, as shown in Fig.
537 [8A](#), can thus result in overestimations in the derived monthly CO emission estimates because
538 the assimilation system will tend to adjust emissions to reduce the initial condition-induced
539 biases.

540 Following Jiang et al. (2017), a suboptimal sequential Kalman filter (Todling and Cohn,
541 1994; Tang et al., 2022) was employed in this work to optimize the modeled CO concentrations
542 with an hourly resolution by combining GC-Adjoint-HEMCO forward simulation and
543 MOPITT CO observations. The CO concentrations provided by the Kalman filter assimilations
544 were archived at the beginning of each month, which were used as the optimized monthly initial
545 CO conditions in the inverse analysis. As shown in Fig. [8B](#), the biases in the modeled CO
546 columns in the optimized initial CO conditions are pronounced lower than those in the original
547 simulation (Fig. [8A](#)). The optimization of the initial CO conditions is essential for our inverse

Deleted: 7A

Deleted: the model

Deleted: 7A

Deleted: 7B

Deleted: 7A

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

556 Fig. 9A shows the distribution of a priori anthropogenic CO emissions in 2015. The
557 regional a priori anthropogenic CO emissions (as shown in Table 4) are 243.53, 34.42, 23.24,
558 30.39, 25.94 and 2.02 Tg/y over Asia, North America, Africa, South America, Europe, and
559 Australia, respectively. As shown in Fig. 9B, our inverse analysis suggests a wide distribution
560 of underestimations in the a priori anthropogenic CO emissions in 2015 except in E. China.
561 The regional scaling factors (Table 4) are 1.16, 1.47, 1.52, 1.41, 1.60 and 1.38, and the a
562 posteriori anthropogenic CO emissions are 283.20, 50.47, 35.34, 42.92, 41.62 and 2.79 Tg/y
563 over Asia, North America, Africa, South America, Europe, and Australia, respectively. As
564 shown in Fig. 9C, we find noticeable underestimations in the modeled CO columns in the a
565 priori simulations, despite the negative biases being much weaker than those in Fig. 8A due to
566 the optimization of the initial CO conditions. The negative biases are effectively reduced in the
567 a posteriori simulation driven by the a posteriori CO emission estimates (Fig. 9D).

568 Finally, we compare the a posteriori CO emission estimates in this work with Jiang et al.
569 (2017), who constrained CO emissions in 2001-2015 with GC-Adjoint-STD by assimilating
570 the same MOPITT CO observations. As shown in Table 4, the a posteriori anthropogenic CO
571 emission estimates in this work match well with Jiang et al. (2017) in North America and Africa
572 but are 38%, 157% and 228% higher than those in Jiang et al. (2017) in Asia, South America
573 and Australia, respectively. A major discrepancy between this work and Jiang et al. (2017) is
574 the treatment of ocean grids. Jiang et al. (2017) defined ocean grids as continental boundary
575 conditions, which were rewritten hourly using the optimized CO concentrations archived from
576 the suboptimal sequential Kalman filter by assimilating MOPITT CO observations. Only
577 MOPITT data over land were assimilated in the 4D-var assimilations in Jiang et al. (2017)

Deleted: 8A

Deleted: 8B

Deleted: 8C

Deleted: 7A

Deleted: 8D

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

588 As shown in Fig. 9D, the a posteriori simulation demonstrates positive biases in CO
589 columns over China and Southeast Asia, which is a signal of overestimated local CO emissions;
590 meanwhile, the negative biases over the northern Pacific Ocean are reduced in the a posteriori
591 simulation. The negative biases over the remote ocean are more affected by CO chemical
592 sources and sinks; however, biases in chemical sources cannot be effectively adjusted because
593 of the global uniform scaling factor for CH₄-generated CO emissions; biases in chemical sinks
594 cannot be adjusted because of the fixed OH fields in the tagged-CO simulation. Jiang et al.
595 (2017) tried to address this problem by defining continental boundary conditions so that the
596 inverse analysis is dominated by local MOPITT observations to avoid the influence of model
597 biases accumulated within the long-range transport. Conversely, CO emissions over China and
598 Southeast Asia are overestimated in this work to offset the negative biases over the northern
599 Pacific Ocean. We expect similar overestimations in the a posteriori CO emission estimates
600 over South America, southern Africa, and Australia in this work because it is the effective
601 pathway to reduce the negative bias over the ocean in the Southern Hemisphere.

602 4. Conclusion

603 This work demonstrates our efforts on the development of a new framework to facilitate
604 emission inventory updates in the adjoint of GEOS-Chem model. The major advantage of this
605 new framework is good readability and extensibility, which allows us to conveniently support
606 HEMCO emission inventories, including CEDS, MIX, NEI2011, DICE_AF, AF_EDGAR43,
607 APEI and GFED4. The updated emission inventories are critical for reliable sensitivity

Deleted: 8D

Deleted: An updated version (GC-Adjoint-HEMCO)

Deleted: was developed in this work.

Deleted: updates

Deleted: GC-Adjoint-HEMCO include 1)

Deleted: for the MERRA-2 meteorological data so that we can perform long-term inverse analysis with consistent meteorological data in 1979-present and 2) support for the HEMCO including

Deleted: inverse analysis

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

625 Two types of OSSE were conducted to evaluate the model performance in 4D-var
626 assimilations. The a posteriori CO emissions converged to the true states in all major emission
627 regions with fully covered pseudo-CO observations; the limited coverage of observations by
628 sampling the pseudo-CO observations at the locations/times of MOPITT CO observations and
629 smoothing with MOPITT averaging kernels resulted in approximately 15% underestimations
630 in the a posteriori CO emissions over North America and Europe. Furthermore, as an example
631 application of the developed capabilities, we constrain anthropogenic CO emissions in 2015
632 by assimilating MOPITT CO observations. The a posteriori anthropogenic CO emission
633 estimates derived in this work match well with Jiang et al. (2017) in North America and Africa
634 but are overestimated in Asia, South America and Australia, which could be associated with
635 the different treatment of MOPITT CO observations over ocean grids and the large differences
636 in CO chemical sources and sinks. The capabilities developed in this work are a useful
637 extension for the adjoint of GEOS-Chem model. More efforts are needed to support emissions
638 inventories associated with full chemistry simulations, as well as integration of these
639 capabilities with the standard GEOS-Chem adjoint code base for better development of the
640 community of the adjoint of GEOS-Chem model.

Deleted: they are helpful for

Deleted: 4D-var

Deleted: the

Deleted: from various inventories, chemical sources and sinks, and

Deleted: provided by GC-Adjoint-HEMCO are compared with those provided by GC-v12 and GC-Adjoint-STD, and we find good agreement

Deleted: the

Deleted: between GC-Adjoint-HEMCO and GC-v12.

Deleted: are employed

Deleted: of GC-Adjoint-HEMCO

Deleted: converge

Deleted: GC-Adjoint-HEMCO

Deleted: GC-Adjoint-HEMCO

Deleted: is

Deleted: tagged-CO simulation

Deleted: higher spatial resolutions and to support

Deleted: simulation in GC-Adjoint-HEMCO.

661 **Code and data availability:** The MOPITT CO data can be downloaded from
662 <https://asdc.larc.nasa.gov/data/MOPITT/>. The GEOS-Chem model (version 12.8.1) can be
663 downloaded from http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_12#12.8.1.
664 The adjoint of GEOS-Chem model (GC-Adjoint-STD) can be downloaded from
665 http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint. The adjoint of
666 GEOS-Chem model (GC-Adjoint-HEMCO) can be downloaded from
667 <https://doi.org/10.5281/zenodo.7512111>.

668

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

672

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

674

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

680

681 **Tables and Figures**

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

683

684 **Table 2.** Regional combustion CO emissions, VOC-generated CO (PCO_NMVOC), CH₄-
685 generated CO (PCO_CH₄), CO sinks (CO_OH, calculated as CO_OH = KRATE×CO×OH),
686 and simulated surface and column CO concentrations in 2015. The region definitions are shown
687 in Fig. 2A.

688

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

695

696 **Table 4.** Regional anthropogenic CO emissions (with unit Tg/y) and annual scaling factors in
697 2015 in this work and Jiang et al. 2017.

698

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

700

701 **Fig. 2.** Total combustion CO emissions in 2015 from (a) GC-v12; (b) GC-Adjoint-HEMCO;
702 (c) GC-Adjoint-STD. The unit is molec/cm²/s.

703

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

706

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

711

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

715

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

720

Deleted: .

Deleted: .

Deleted: .

Deleted: ¶

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

Moved up [2]: Fig.

Deleted: 6

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

Deleted: 7

734 **Fig. 9.** (a) A priori anthropogenic CO emissions in 2015 with unit molec/cm²/s; (b) Annual
735 scaling factors for CO emissions in 2015. The scaling factors represent the ratio of the estimated
736 to true emissions. (c-d) the a priori and a posteriori biases calculated by (model-
737 MOPITT)/MOPITT.

Deleted: 8

738

739 References

- 740 Barré, J., Edwards, D., Worden, H., Da Silva, A., and Lahoz, W.: On the feasibility of
741 monitoring carbon monoxide in the lower troposphere from a constellation of Northern
742 Hemisphere geostationary satellites. (Part 1), *Atmos Environ*, 113, 63-77,
743 10.1016/j.atmosenv.2015.04.069, 2015.
- 744 Dedoussi, I. C., Eastham, S. D., Monier, E., and Barrett, S. R. H.: Premature mortality related
745 to United States cross-state air pollution, *Nature*, 578, 261-265, 10.1038/s41586-020-1983-8,
746 2020.
- 747 Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martínez-Alonso, S., Worden, H.
748 M., and Sweeney, C.: A climate-scale satellite record for carbon monoxide: the MOPITT
749 Version 7 product, *Atmos Meas Tech*, 10, 2533-2555, 10.5194/amt-10-2533-2017, 2017.
- 750 Fisher, J. A., Murray, L. T., Jones, D. B. A., and Deutscher, N. M.: Improved method for linear
751 carbon monoxide simulation and source attribution in atmospheric chemistry models
752 illustrated using GEOS-Chem v9, *Geosci Model Dev*, 10, 4129-4144, 10.5194/gmd-10-4129-
753 2017, 2017.
- 754 Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of
755 global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and
756 Aerosols from Nature), *Atmos Chem Phys*, 6, 3181-3210, 10.5194/acp-6-3181-2006, 2006.
- 757 Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R.
758 C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., Brauer, M., Apte, J. S., Henze, D. K.,
759 Zhang, L., Zhang, Q., Ford, B., Pierce, J. R., and Martin, R. V.: Global Estimates and Long-
760 Term Trends of Fine Particulate Matter Concentrations (1998-2018), *Environ Sci Technol*,
761 54, 7879-7890, 10.1021/acs.est.0c01764, 2020.

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

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

772 Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T.,
773 Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N.,
774 Kurokawa, J.-i., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.:
775 Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the
776 Community Emissions Data System (CEDS), *Geosci Model Dev*, 11, 369-408, 10.5194/gmd-
777 11-369-2018, 2018.

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

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

785 Jiang, Z., Worden, J. R., Jones, D. B. A., Lin, J. T., Verstraeten, W. W., and Henze, D. K.:
786 Constraints on Asian ozone using Aura TES, OMI and Terra MOPITT, *Atmos Chem Phys*,
787 15, 99-112, 10.5194/acp-15-99-2015, 2015b.

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

791 Jiang, Z., Zhu, R., Miyazaki, K., McDonald, B. C., Klimont, Z., Zheng, B., Boersma, K. F.,
792 Zhang, Q., Worden, H., Worden, J. R., Henze, D. K., Jones, D. B. A., Denier van der Gon,
793 H. A. C., and Eskes, H.: Decadal Variabilities in Tropospheric Nitrogen Oxides Over United
794 States, Europe, and China, *J Geophys Res-Atmos*, 127, e2021JD035872,
795 10.1029/2021jd035872, 2022.

796 Jones, D. B. A., Bowman, K. W., Palmer, P. I., Worden, J. R., Jacob, D. J., Hoffman, R. N.,
797 Bey, I., and Yantosca, R. M.: Potential of observations from the Tropospheric Emission
798 Spectrometer to constrain continental sources of carbon monoxide, *J Geophys Res-Atmos*,
799 108, 2003JD003702, 10.1029/2003jd003702, 2003.

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

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

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

810 Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., and Bates, K. H.: Anthropogenic drivers of
811 2013-2017 trends in summer surface ozone in China, *Proc Natl Acad Sci USA*, 116, 422-427,
812 10.1073/pnas.1812168116, 2019.

813 Li, M., Zhang, Q., Kurokawa, J.-i., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D.
814 G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and
815 Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international
816 collaboration framework of the MICS-Asia and HTAP, *Atmos Chem Phys*, 17, 935-963,
817 10.5194/acp-17-935-2017, 2017.

818 Lin, H., Jacob, D. J., Lundgren, E. W., Sulprizio, M. P., Keller, C. A., Fritz, T. M., Eastham,
819 S. D., Emmons, L. K., Campbell, P. C., Baker, B., Saylor, R. D., and Montuoro, R.:
820 Harmonized Emissions Component (HEMCO) 3.0 as a versatile emissions component for
821 atmospheric models: application in the GEOS-Chem, NASA GEOS, WRF-GC, CESM2,
822 NOAA GEFS-Aerosol, and NOAA UFS models, *Geosci Model Dev*, 14, 5487-5506,
823 10.5194/gmd-14-5487-2021, 2021.

824 Qu, Z., Henze, D. K., Worden, H. M., Jiang, Z., Gaubert, B., Theys, N., and Wang, W.:
825 Sector - Based Top - Down Estimates of NO_x, SO₂, and CO Emissions in East Asia, *Geophys*
826 *Res Lett*, 49, e2021GL096009, 10.1029/2021gl096009, 2022.

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

832 Tang, Z., Chen, J., and Jiang, Z.: Discrepancy in assimilated atmospheric CO over East Asia
833 in 2015–2020 by assimilating satellite and surface CO measurements, *Atmos Chem Phys*, 22,
834 7815-7826, 10.5194/acp-22-7815-2022, 2022.

835 Todling, R., and Cohn, S. E.: Suboptimal schemes for atmospheric data assimilation based on
836 the Kalman filter, *Monthly Weather Review*, 122, 10.1175/1520-
837 0493(1994)122<2530:SSFADA>2.0.CO;2, 1994.

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

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

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

849 Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P.:
850 Observations of near-surface carbon monoxide from space using MOPITT multispectral
851 retrievals, *Journal of Geophysical Research*, 115, D18314, 10.1029/2010jd014242, 2010.
852 Zhang, L., Chen, Y., Zhao, Y., Henze, D. K., Zhu, L., Song, Y., Paulot, F., Liu, X., Pan, Y.,
853 Lin, Y., and Huang, B.: Agricultural ammonia emissions in China: reconciling bottom-up and
854 top-down estimates, *Atmos Chem Phys*, 18, 339-355, 10.5194/acp-18-339-2018, 2018.
855 Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z.,
856 Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.:
857 Asian emissions in 2006 for the NASA INTEX-B mission, *Atmos Chem Phys*, 9, 5131-5153,
858 10.5194/acp-9-5131-2009, 2009.
859 Zhao, H., Geng, G., Zhang, Q., Davis, S. J., Li, X., Liu, Y., Peng, L., Li, M., Zheng, B., Huo,
860 H., Zhang, L., Henze, D. K., Mi, Z., Liu, Z., Guan, D., and He, K.: Inequality of household
861 consumption and air pollution-related deaths in China, *Nat Commun*, 10, 4337,
862 10.1038/s41467-019-12254-x, 2019.
863 Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines
864 for Large-Scale Bound Constrained Optimization, *ACM Transactions on Mathematical*
865 *Software*, 23, 550-560, 10.1145/279232.279236, 1997.
866