Concentration Trajectory Route of Air pollution with an Integrated Lagrangian Model (C-TRAIL Model v1.0) Derived from the Community Multiscale Air Quality Model (CMAQ Model v5.2)

1   Arman Pouyaei [1], Yunsoo Choi [1], Jia Jung [1], Bavand Sadeghi [1], Chul Han Song[2]

2   [1] Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA
3   [2] School of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, South
4   Korea

     Correspondence to: Yunsoo Choi (ychoi6@uh.edu)

Abstract. This paper introduces a novel Lagrangian model (Concentration Trajectory Route of Air pollution with an Integrated Lagrangian model, C-TRAIL version 1.0) output from a Eulerian air quality model for validating the source-receptor direct link by following polluted air masses. To investigate the concentrations and trajectories of air masses simultaneously, we implement the trajectory-grid (TG) Lagrangian advection scheme in the CMAQ (Community Multiscale Air Quality) Eulerian model version 5.2. The TG algorithm follows the concentrations of representative air "packets" of species along trajectories determined by the wind field. The diagnostic output from C-TRAIL accurately identifies the origins of pollutants. For validation, we analyze the results of C-TRAIL during the KORUS-AQ campaign over South Korea. Initially, we implement C-TRAIL in a simulation of CO concentrations with an emphasis on the long- and short-range transport effects. The output from C-TRAIL reveals that local trajectories were responsible for CO concentrations over Seoul during the stagnant period (May 17-22, 2016) and during the extreme pollution period (May 25-28, 2016), highly polluted air masses from China were distinguished as sources of CO transported to the Seoul Metropolitan Area (SMA). We conclude that during the study period, long-range transport played a crucial role in high CO concentrations over the receptor area. Furthermore, for May 2016, we find that the potential sources of CO over the SMA were the result of either local transport or long-range transport from the Shandong Peninsula and, in some cases, from regions north of the SMA. By identifying the trajectories of CO concentrations, one can use the results from C-TRAIL to directly link strong potential sources of pollutants to a receptor in specific regions during various time frames.

Keywords: C-TRAIL, Trajectory analysis, CMAQ, East Asia, KORUS-AQ campaign

1   Introduction

Determining the long-range transport (LRT) of pollutants has been a challenge for air quality researchers. As the chemical composition of outflow over a region or continent can significantly affect air quality downwind, information about LRT must be reliable. Several studies have applied a number of methods to examine the role that LRT plays in the concentrations of particulate matter (PM), ozone, trace gases, and biomass burning tracers over target regions (Stohl, 2002). For instance, in an attempt to identify possible sources of PM in East Asia, several studies (Choi et al., 2014; Lee et al., 2019; Oh et al., 2015; Pu et al., 2015) have applied the NOAA Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler, 1998) and back-trajectory analysis. A number of researchers have incorporated the widely used HYSPLIT model into other chemical-transport models (CTMs) to measure the LRT of ozone, carbon monoxide (CO), and aerosols to establish the source-receptor relationship of air masses over the United States (Bertschi and Jaffe, 2005; Carroll et al., 2008; Gratz et al., 2015; Price et al., 2004; Sadeghi et al., 2020; Weiss-Penzias et al., 2004). Several studies have used another model, the FLEXTRA trajectory model (Stohl, 1996; Stohl and Seibert, 1998), to capture the background source regions of high-PM over East Asia and quantify the contributions from these regions (Lee et al., 2011, 2013). Furthermore, this model has also been applied to some European regions to explain the potential advected contribution of aerosols (Cristofanelli et al., 2007; Petetin et al., 2014; Salvador et al., 2008). Several studies have recently attempted to develop new trajectory models that overcome truncation errors that originate from schemes for numerically integrating trajectory equations (Döös et al., 2017; Rößler et al., 2018) and to link trajectories to specific trace species (Kruse et al., 2018; Stenke et al., 2009). Another widely used tool for studying the distribution of CO, ozone, PM, and other aerosols for both air quality forecasting and emission scenario analysis is the EPA Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006). CMAQ, supported by meteorological inputs from the Weather Research and Forecasting (WRF) model, or advanced machine learning-based methods (Eslami et al., 2019; Lops et al., 2019; Sayeed et al., 2020), assists policy-makers with solving pollution-related issues by legislating regulations. Spatial concentration patterns of pollutants incorporated with other models (i.e., back-trajectory models) or satellite data enhance our understanding of the impact of LRT and other related processes such as the formation of aerosols, emissions, and dry deposition in various regions (Chen et al., 2014; Chuang et al., 2008, 2018; Souri et al., 2016; Wang et al., 2010; Xu et al., 2019; Zhang et al., 2019).

The conventional way of estimating potential source regions of air-mass transport is to use back-trajectory modeling. Frequently used for source-receptor linkage, such models combine their output with measurements of pollutant concentrations. As this source-receptor linkage approach uses meteorology-based models for back trajectories, it is not fully accepted because it is unable to directly determine whether an originated air mass is polluted or non-polluted (Lee et al., 2019). Thus, back-trajectory modeling sometimes provides unreliable information from which to assess the variation of pollutants at a receptor point, raising concern about its use for interpreting the contribution of the effect of LRT on concentrations of a target pollutant. In addition, other factors such as emissions and the local production of air pollutants contribute to variation in a target pollutant.

Although aircraft campaigns in several regions have applied a Lagrangian approach for interpreting variations in concentrations, they have not effectively addressed the above concern. After all, such campaigns are neither frequent nor continuous.

In this study, we implement a Lagrangian advection scheme that we refer to as the trajectory grid (TG) (Chock et al., 1996), into the Eulerian CMAQ v5.2 model. We introduce a new type of output from the Concentration Trajectory Route of Air pollution with the Integrated Lagrangian (C-TRAIL v1.0) stand-alone model in addition to CMAQ v5.2 output to simultaneously accomplish two objectives: (1) to provide a direct link between polluted air masses from sources and a receptor and (2) to provide the spatial concentration distribution of several pollutants that explains relevant physical processes. Chock et al. (2005) incorporated the TG into an air quality model to study the accuracy of this Lagrangian advection method over the Bott advection scheme applied in the Eulerian domain. One significant outcome of the TG model applied to CTMs is its ability to account for the concentrations of pollutants in air masses in its investigation of trajectories. This outcome addresses the unreliability of meteorology-based Lagrangian models when the pollutedness or cleanliness of an originated air mass becomes an issue. For this study, we have selected CO as our trace gas target. As this pollutant has an oxidation lifetime of approximately two months, it is an ideal tracer with which we can study its impact on LRT without having stable background levels such as CO_2 (Heald et al., 2003; Liu et al., 2010; Vay et al., 2011). Furthermore, as CO is produced mainly by the incomplete combustion of carbon-containing fuels (Halliday et al., 2019), it is an ideal proxy with which we can relate concentrations of receptors to sources of traffic or power-plant emissions. We begin by introducing the methodology behind TG and the implementation of TG into CMAQ. Then, we present a simple case and our interpretation of the C-TRAIL output. Finally, we present a case study of C-TRAIL for Korea and the United States Air Quality (KORUS-AQ) campaign over South Korea.

2   Methodology

2.1   Description of the TG approach

To solve the transport equation, Chock et al. (1996) presented the TG approach in air quality modeling. This approach, which entails transporting points on a concentration profile along their trajectories in a Lagrangian manner, uses the Eulerian approach for diffusive transport. From now on, we will refer to these points as "air packets" for two reasons: (1) Their nature is similar to that of air parcels, but they are massless, and (2) they behave much like particles, but they carry information about several species. The TG method rewrites the advection equation for concentration as follows:

$$\frac{dC}{dt} = \frac{\partial C}{\partial t} + \mathrm{v} \cdot \nabla C = -(\nabla \cdot \mathrm{v})C, \qquad\qquad (1)$$

86   where C is the concentration of species in velocity field v. The Lagrangian approach divides the total derivative of the

87   concentration into a full derivative of concentration with respect to time, $\frac{dC}{dt}$, and a remaining term containing velocity

88   divergence, $-(\nabla \cdot v)C$. Following this approach, the TG automatically and accurately conserves the mass, sign, and shape of

89   the concentration profile. As interpreted from the equation, the concentration profile of the species along trajectories can be

90   described. Otherwise stated, after determining the location of a packet and the concentration inside the domain, we are able to

91   assess the concentration profile along its trajectory. Since all species represented in one packet and all of the packets move in

92   the flow field according to the wind velocity, differentiating among advection equations for each species (as is done in Eulerian

93   advection schemes) is no longer necessary; thus, this approach removes the associated numerical errors with the discretization

94   of the advection equation. The concentration of each packet along its trajectory can be determined by the following equation:

95

$$C(t) = C(t_0)\, exp\left(-\int_{t_0}^{t}(\nabla \cdot v)dt\right) \approx C(t_0)\, exp[-(\nabla \cdot v)(t - t_0)], \tag{2}$$

96

97   where $C(t)$ is the concentration of species at the location of a packet as it moves along its trajectory. Since we can use the TG

98   method to calculate the concentration from an ordinary differential equation, it is mass conserving, monotonic, and accurate.

99   Although interpolation errors occur during the diffusion step, they are typically considerably smaller than Eulerian advection

100   errors (Chock et al., 2005). In addition, the trajectory will be three-dimensional and as accurate as the input for wind velocity

101   and direction. In particular, for large-scale vertical winds, in which CTMs typically modify the scheme to address the mass-

102   conservation issue, TG removes numerical diffusion from upwind vertical advection schemes and generates more physical

103   vertical winds (Hu and Talat Odman, 2008). It is worth mentioning that units for the concentration of species are referred to

104   as ppbv or μgm⁻³, depending on the species type, and the unit conversion is taken into account in the process of solving

105   equations.

2.2   Implementation of TG in CMAQ v5.2

107   In this section, we briefly describe the key features of TG implementation in the CMAQ v5.2 model, a Eulerian model

108   consisting of several modules (i.e., advection, diffusion, cloud, and aqueous-phase). The C-TRAIL v1.0 model requires the

109   same meteorology, initial conditions (ICs), boundary conditions (BCs), and emissions as CMAQ. All CMAQ modules and

110   parameters are associated with cells of the Eulerian grid in the model domain. Since TG is based on CMAQ in this study and

111   some of the CMAQ processes cannot be satisfactorily carried out by Lagrangian models (e.g., eddy diffusion) at this time, grid

112   cells are the primary structure for initiating and listing packets. By grouping the packets into grid cells, keeping track of which

113   packets are close to each other is easier. While the grid cells of Eulerian models represent Eulerian-type outputs, tracking the

114   packets of Lagrangian advection provides both their trajectories and their concentrations (Figure 1).

115

116   The process of advection for packets follows the ordinary differential equation:

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{V}(\mathbf{y}(t), t), \tag{3}$$

117 where $\mathbf{V}$ [ms^{-1}] is the three-dimensional wind velocity and $\mathbf{y}(t)$ [m] is the position vector of packets at time t [s]. The equation
118 is solved using the following simple predictor-corrector scheme:

$$\mathbf{y}^i(t + \Delta t) = \mathbf{y}(t) + \mathbf{V}(\mathbf{y}(t), t)\Delta t \tag{4}$$

$$\mathbf{y}^f(t + \Delta t) = \mathbf{y}(t) + 0.5[\mathbf{V}(\mathbf{y}(t), t) + \mathbf{V}(\mathbf{y}^i(t + \Delta t), t + \Delta t)]\Delta t, \tag{5}$$

119 where $\mathbf{y}^i$ is the initial estimate of the new position from the predictor step and $\mathbf{y}^f$ is the final position calculated by the
120 corrector step. When the initiated packets in the domain follow the Lagrangian equation, they land in different grid cells after
121 each time step. To balance the density of packets in grid cells, we apply a simple packet management technique that includes
122 spawning (filling) and pruning (emptying) processes. In the spawning process, every step entails the creation of a group of
123 new packets in each cell with insufficient packets. The initial composition of a spawned packet is estimated from nearby
124 packets. The pruning process entails the removal of extra packets from cells that have become overpopulated. During this
125 process, the packets closest to the cell center are retained. Such packet management with favorable options contributes to
126 reducing the computational costs of the C-TRAIL model. The limitation of this packet management approach, however, is that
127 it violates mass conservation. These errors are caused by sub-grid interpolations of packets in the spawning or pruning process.
128 The underlying algorithms for both vertical and horizontal diffusion, emissions, and other processes are the same as those in
129 standard CMAQ (Byun and Schere, 2006) with some minor modifications. The coupling of Eulerian diffusion and TG
130 advection at each time step is accomplished by first taking the average of concentrations from all packets in each cell as the
131 cell average. Then, by considering each packet as a cell and cell average representative of neighboring cells, we use a predictor-
132 corrector method to determine the concentration of each packet. In addition, the C-TRAIL model considers convective
133 transport only for resolved clouds (when clouds cover an entire grid). The WRF model implements the cloud model to obtain
134 cloud properties on a sub-grid scale and addresses vertical transport on a resolved scale (Kain, 2004). Outputs from the WRF
135 model are used in the CMAQ's cloud model to account for convective transport in two separate modules: sub-grid scale clouds
136 and resolved clouds (Byun and Schere, 2006). In this version of C-TRAIL for convection, we only use vertical winds
137 determined from resolved clouds (see Table S1). Figure 2 summarizes the process of C-TRAIL from initialization to output
138 generation. By combining the locations of the packets from each time step during the 24 hours, we generate the 24-hour
139 trajectory of each packet.

3   Setup and Validation of the Model

141 In this study, we implement TG in the CMAQ model version 5.2. Shown in Figure 3, the model domain, with a horizontal grid
142 resolution of 27-km over East Asia, covers the eastern parts of China, the Korean Peninsula, and Japan. We use the 2010 MIX
143 emission inventory (Li et al., 2017) at a 0.25-degree spatial resolution. The emission inventory contains monthly averaged
144 carbon bond version 5 (Sarwar et al., 2012) emission information, which includes ten chemical species, including CO, in five
145 different sectors. We also use the 2011 Clean Air Policy Support System emission high-resolution (1-km) inventory from the

National Institute of Environmental Research for Korea, which contains the area and the line and point sources of a variety of species, including CO. We provide WRF model v3.8 output as the meteorological input in our CMAQ model. We validate the wind predictions from our WRF model with surface measurements and radiosonde measurements from the KORUS-AQ period (see Table S2-S3 and Figure S1-S4). Jung et al. (2019) validated the air quality model set up by comparing simulated and observed aerosol optical depths; they showed a correlation of 0.64 for the entire KORUS-AQ campaign period. Their comparison of various gaseous and particulate species also showed close agreement with observations.

We run C-TRAIL simulations for May 2016 during the KORUS-AQ campaign. Studies pertaining to this campaign (Al-Saadi et al., 2016; Choi et al., 2019; Miyazaki et al., 2019) have separated the time frame into three periods (Table 1) based on meteorological conditions: 1) the dynamic weather period (DWP), a rapid cycle of clear and rainy days in the Korean Peninsula (May 10-16); 2) the stagnant period (SP), in which the area was under the influence of a high-pressure system (May 17-22) and which showed the influence of local emissions; and 3) the extreme pollution period (EPP) with high peaks of pollutants that showed strong direct transport from China (May 25-28).

The overall accuracy of the CMAQ CO simulation compared to aircraft measurements during all periods is presented in Figure 4(a). The correlation between the modeled CO concentrations and observations at various altitudes for the entire month of May 2016 was 0.71, indicating that the performance of the model is sufficiently reliable for a study of the sources of CO concentration (Table 1). Figure 4 illustrates the under-prediction of the model during the DWP and SP. However, the model shows a high correlation during the EPP compared to higher CO observations over the Korean Peninsula. We also provide a CMAQ CO comparison with surface station measurements in the supplementary document (see Table S4 and Figure S5). The results of this comparison also show the under-prediction of the model, caused by uncertain emission inventories over East Asia. The C-TRAIL outputs of the mentioned periods will be discussed in Section 3.3.

The Eulerian output from CMAQ, including CO concentrations and surface wind fields, is displayed in Figure 5. High peaks of CO concentrations appeared in south-eastern China, including the Shanghai region and the Shandong Peninsula, because of high anthropogenic emissions in these areas. The impact on pollution from LRT was greater in this region because the dominant wind over East Asia in May was westerly, which explains our observations of high CO concentrations over the Yellow Sea. We also observed a shallow anticyclone (a common phenomenon that affects the regional transport of pollution in this region) over the Yellow Sea during the study period. From a thorough investigation of CO concentrations and wind patterns during various meteorological periods, we present the following major findings: 1) During the DWP, a mixed response from the LRT of CO and local emissions occurred. Also, considering the impact of convection, the concentrations of CO over Korea could have been increased or decreased because of vertical wind transport and cloud updrafts and downdrafts. Owing to the dynamic nature of this period (i.e., cloudy, rainy, or clear), the interpretation of the LRT effect by conventional methods poses a challenge. 2) During the SP, a high-pressure system settled over the Korean Peninsula, which explains the extremely low wind

speed and the stagnant air, the latter of which eliminated the impact of LRT. Even though one might assume that the model would produce more accurate simulations with less convection-related transport, CO concentrations were significantly underestimated by the model (Jeon et al., 2016) because of uncertainties in the chemistry modeling and the faulty emission inventories over East Asia. 3) During EPP, as shown in Figure 5, the anticyclone over the Yellow Sea contributed to the transport of more CO from China to the Korean Peninsula. We also observed high concentrations of CO in regions throughout China were. Thus, the combination of these two effects produced model predictions of higher concentrations over Korea.

The raw hypothesis from Eulerian outputs is that a high CO concentration at a receptor during a specific period is due to LRT from a source because the direction of the wind is typically toward the receptor during the period of simulation. This hypothesis is based on the average wind speed and direction and the average CO concentration, which do not constitute a reliable source of this assumption. We will briefly explain why we require merged output with simultaneous changes in trajectories and concentrations. To determine the source of LRT, researchers should include one major parameter in their investigations: the trajectory of the air mass. Once the location of the source and the trajectory of the air mass are known, the air mass is assumed to be polluted. If the air mass is not polluted, then that source is not responsible for high concentrations in the receptor location. Therefore, linking the source to the receptor based on only mean wind patterns and concentrations is not a reliable approach. The following section will discuss how we combine concentrations and trajectories into one set of outputs to explain the trajectories more clearly.

4   Analysis of C-TRAIL

Because C-TRAIL is a diagnostic tool derived from CMAQ, both a Lagrangian output and CMAQ standard Eulerian output are available after each run. C-TRAIL helps us identify the source-receptor linkage, save the full trajectory of packets, and display the path of selected packets. Therefore, not only do C-TRAIL simulations provide all spatial concentration changes, but they also display the trajectories of each packet, owing to the Lagrangian approach of TG. In addition, we are able to determine changes in concentrations along this trajectory. The difference between this model and other meteorological-based models is that they enable us to study changes in the concentrations of selected species along different paths, investigate evidence for the amount of pollution in originated air masses, study the reason behind the oscillation of concentrations, and examine the linkage of oscillations to both sources and sinks along the path.

This section provides an example of how we use C-TRAIL to study the sources of different packets from different altitudes (from below 1 km to almost 10 km) over the Seoul Metropolitan Area (SMA); later sections will focus on the entire month of May 2016 C-TRAIL over the SMA and provide more comprehensive illustrations of the concentrations and altitudes of trajectories. Figure 6 displays the C-TRAIL output for June 4, 2016. We gathered all of the packets over the city of Seoul and analyzed the trajectory of each packet. Figure 6(a) shows the path of all the packets, represented by various colors, from their

212 sources. We observed that some of the packets came from south-eastern South Korea, and one originated in south-eastern
213 China, traveled over the Yellow Sea, and landed in Seoul. Some of the packets also originated from northwest of South Korea
214 and northern China. Most of the packets, however, were locally initiated, generally from regions around the SMA. Using the
215 HYSPLIT back-trajectory model, we found relatively similar trajectories (Figure S6).
216
217 Figure 6(b) depicts how the CO concentrations of the four most aged packets changed as they traveled on their path toward
218 Seoul. This type of output is a new feature that has not been studied before. With meteorological-based back-trajectory models,
219 the path of air parcels and their back trajectories can be delineated; we are the first, however, to use a CMAQ-based Lagrangian
220 integrated model to study the concentrations of species (in this case, CO) via the paths of air packets. The four most aged
221 packets came from 24, 21, 20, and 17 time-steps back (Figure 6(b)). We find these packets interesting because they follow a
222 long path, changes in their concentrations fluctuate, and they are easy to comprehend. From studying these packets and their
223 C-TRAILs, we generally understand that the concentration of each packet increases as it approaches the SMA. The
224 concentrations of near-surface packets tend to fluctuate more than those of high-altitude packets (Figure S7). Also, larger
225 oscillations in the concentrations occur over land rather than over the ocean, which, however, becomes more vivid when a
226 near-surface packet reaches land from the ocean and suddenly peaks in concentration. The sudden peaks in the concentrations
227 of near-surface packets are due to their movement over either a city or some source of emissions. Over the SMA and other
228 cities, two peaks, mainly caused by on-road traffic emissions, occur during local morning and evening times.

229 **5   Case Study for the C-TRAIL Analysis: The May 2016 KORUS-AQ Period**

230 Using a conventional method with model data gathered over the course of a month or a year to incorporate concentrations into
231 a trajectory analysis produces a tremendous amount of outputs that are difficult to interpret simultaneously. For our case study,
232 covering May 2016, we selected Seoul, South Korea, over East Asia as the receptor. We plotted C-TRAIL outputs according
233 to variations in the packet concentrations and their distances from the receptor. Figure 7(a) presents the general path of all
234 packet trajectories reaching the Seoul area at various altitudes at 9:00 AM local time throughout May 2016. The color bar
235 represents the altitude at which the packets were traveling. Generally, packets at low altitudes traveled from local areas to
236 Seoul, and those at high altitudes traveled from more distant regions. One exception was packets that originated in the
237 Shandong Peninsula; Some traveled at high altitudes and some at low altitudes. Figure 7(b) displays a C-TRAIL that
238 represented a unique type of packet that followed the concentrations of trajectories. In this case, each packet at each location
239 (or hour of the trajectory) had a specific CO concentration that depended on its altitude (high altitude/surface), its location
240 (land/sea/urban/forest), and the hour of the day (traffic hours/non-traffic hours). To more clearly explain the location of packets
241 and the variability in their trajectory paths before reaching Seoul, we created a boxplot of packet distances in kilometers from
242 the receptor at each hour before the packets reached Seoul, shown in Figure 7(c). When the packets reached Seoul at 9:00 AM
243 local time, the distance became zero. Furthermore, the boxplot of trajectory heights for all periods is presented in Figure S8.

244   In a study of C-TRAIL outputs, it is better to account for trajectories, concentrations, and distances simultaneously. As a result,

245   the concentrations and distances of packets in early hours (10:00 AM to 2:00 PM of local time) in Figure 7 show high variability

246   in concentrations with a median of around 150 ppbv and a maximum as high as 500 ppbv. Most of these packets originated far

247   from the receptor (i.e., eastern, northern, and south-eastern China). The median concentration, shown in the boxplot, rose

248   slightly between 6:00 PM and 10:00 PM. Distances also showed more variation during this time, which could be explained

249   by the different paths of the trajectories (i.e., local trajectories with shorter distances and LRT trajectories with longer

250   distances). As the packets approached Seoul (6:00 AM to 9:00 AM), the upper whisker of concentration values increased to

251   as high as 400 ppbv, and the distances approached zero, indicating higher concentrations of CO of local trajectories resulting

252   from surface on-road emissions and other emission sources.

253

254   Because of variable weather and wind (i.e., cloudy, rainy, or clear) during the DWP, C-TRAIL showed a mixed response of

255   trajectories from both local and long-range transport, shown in Figure 8(a). A wide interquartile range and a median of close

256   to the 25th percentile at 11:00 AM and 12:00 PM indicate that a few packets contained high concentrations of CO (close to

257   300 ppbv), but the majority consisted of low concentrations (around 100 ppbv). The distance output of low-concentration

258   packets showed distances as long as 500 km (over the Shandong Peninsula). As the packets approached Seoul, the median

259   concentration values were as high as 150 ppbv. Thus, from Figure 8, we conclude that most of the long trajectories followed

260   a path at high altitudes (higher than 7 km), and the polluted trajectories, which originated in the Shandong Peninsula, were

261   from the near-surface, shown in Figure 8(a).

262

263   Unlike the DWP, the SP showed a more vivid display of trajectories, nearly all of which could be considered local trajectories.

264   Long-range trajectories could not be considered responsible for the CO concentration values of Seoul. After all, from 10:00

265   AM to 4:00 PM (

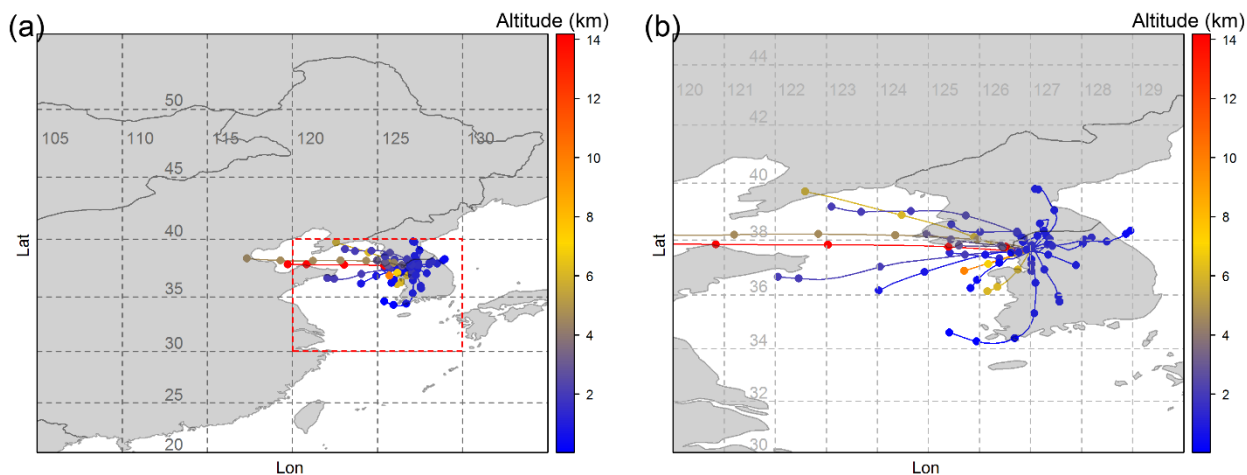
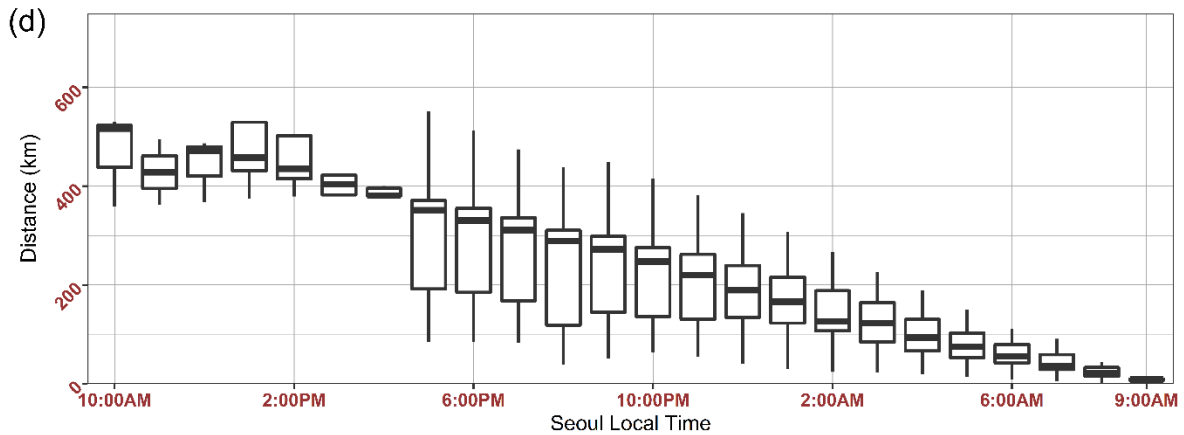
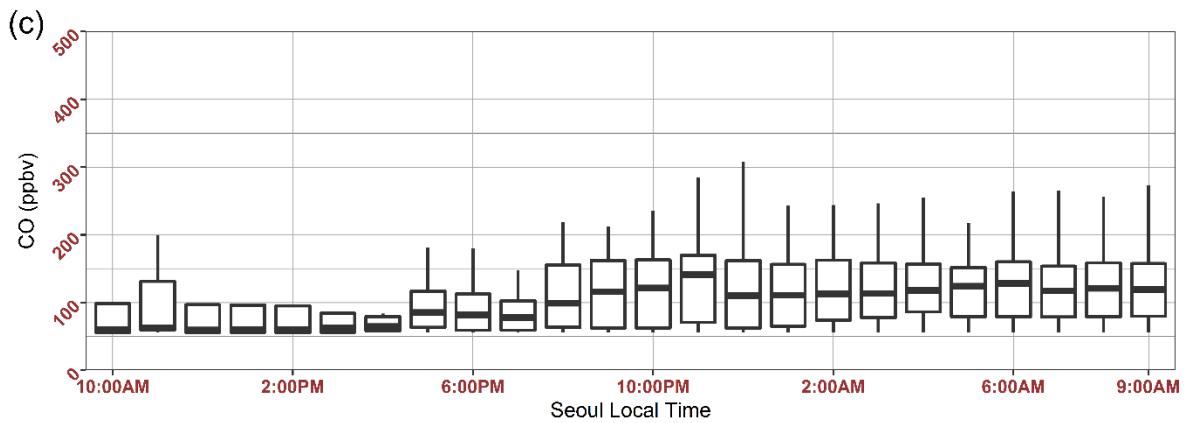



Figure 9(a) and (b)), nearly all of the long-distance packets had concentrations of less than 100 ppbv. The local origination of highly polluted trajectories can be explained by a high-pressure system over the Korean Peninsula during this time period, which was responsible for very low wind speeds. The poor emission inventory over East Asia, however, provided extreme under-predictions of high concentration values during this time period. Therefore, when studying model outputs, we should account for various aspects of the model (e.g., the transport, diffusion, formation, deposition, and convention), in which diffusion, in this case, played a significant role in CO concentration values at the receptor location.

During the EPP, several high concentrations of CO appeared at the early points of trajectories. These high concentrations, combined with high distance values, indicate that the LRT of polluted air masses was responsible for high concentrations of CO during this time period (

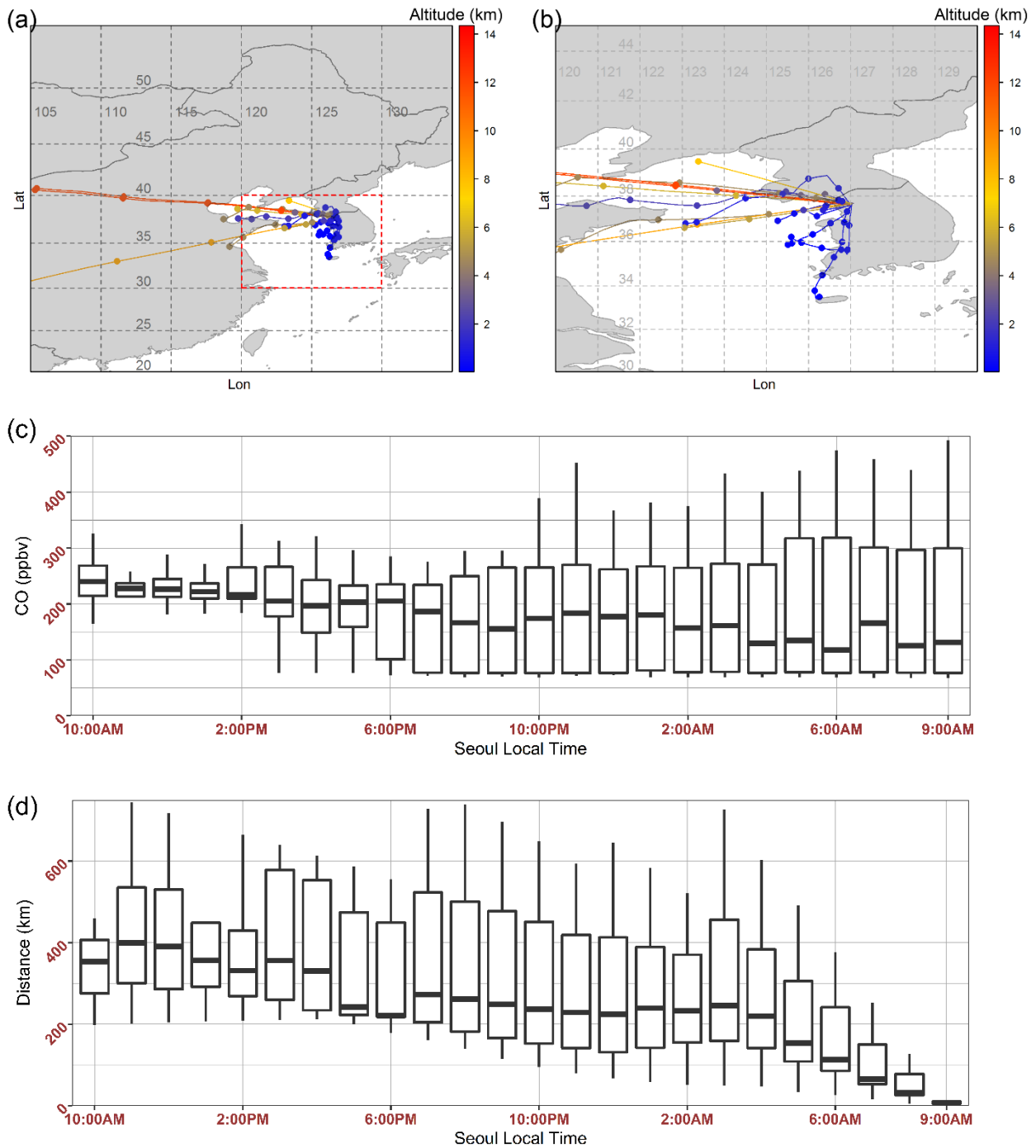



Figure 10). Furthermore, the variability of CO concentrations from 10:00 PM to 9:00 AM at the receptor location stemmed from both the various paths of the trajectories and the distances. High concentration trajectories close to the surface, which originated in the Shandong Peninsula, passed over the Yellow Sea and landed in Seoul at 9:00 AM. When the surface packets

279 reached urban areas, they presented maximum CO concentrations, depending on the time of day and the rush-hour traffic. An

280 assumption made by studies that used Eulerian model outputs or meteorological-based Lagrangian models for this time period

281 was that transport played an important role (Lee et al., 2019). The outputs from C-TRAIL also indicate that highly polluted

282 air masses originated in China (the source) and landed in Seoul (the receptor). That is, the findings of this study regarding the

283 trajectories and the origin of polluted air masses are similar to those of previous studies.

284

285 We further analyzed the diverse aspects of C-TRAIL results using the Open-air package in R (Carslaw and Ropkins, 2012)

286 and determined the frequency of trajectories passing through every one-degree by one-degree gridded area, illustrated in Figure

287 11(a). Central China, northern China, and North Korea were not common areas for packet movement because the packets most

288 likely passed only once through the grids of these regions (at a frequency of about one percent). For the Yellow Sea and the

289 Shandong Peninsula region, however, trajectories more likely passed at a frequency of about ten percent. The figure also

290 shows that most of the trajectories (25 to 100 percent) passed over the west side of the SMA, a two-degree by two-degree area

291 (the dark red section in Figure 11 (a)). We can classify trajectories into separate segments according to their concentrations.

292 Figure 11(b) shows this type of classification and the link between the average concentration of all trajectories to their paths.

293 While higher concentrations were most likely the result of local transport, lower concentrations were most likely from LRT.

294 For the May 2016 case, while most of the high concentration values corresponded to packets that originated in South Korea or

295 close to the SMA, most of the low concentration values corresponded to packets originating in China. Their impact, however,

296 is still evident.

297

298 By clustering the outputs of C-TRAIL, we are better able to locate the dominant paths for the May 2016 trajectories. According

299 to Figure 12(a), based on the Euclidean distance function, about 37.8% of trajectories originated in local areas east, south, and

300 north of the SMA. About 16.1% of trajectories originated in northern China and followed paths over the Yellow Sea to the

301 SMA; about 10.5% of the trajectories came from south-western South Korea and traveled over the Yellow Sea to reach the

302 SMA; about 21.3% of trajectories came from the Shandong Peninsula, and the remaining trajectories (5.3%) originated in

303 central China and were transported over China and the Yellow Sea to the SMA. Angle clustering in Figure 12(b), however,

304 tells a different story about the trajectories. Clustering by the angle distance function shows a similarity among the angles from

305 the starting points of the back trajectories. Generally, nearly all of the packets originated from west of the SMA, 32.2% farther

306 west, 34.5% southwest, 12.7% south/southwest, 14.9% northwest only 5.7% originated from east/southeast of the SMA. This

307 clustering is consistent with strong westerly winds during the spring in East Asia.

308

309 By quantifying clusters based on their trajectories, cluster analysis shows the relative importance of regional sources.

310 Nevertheless, they are not completely accurate at determining the relative contribution of potential source regions because they

311 do not consider concentrations together with trajectories. One method of calculating the probability of potential sources is the

312 potential source contribution function (PSCF), which finds the probability that a source is located at a specific latitude and

313 longitude (Pekney et al., 2006). Figure 12(c) shows that the probability of packets with high concentrations (i.e., those with

314 concentrations at or above 90 percentile) passing over the Yellow Sea and reaching the SMA from the southwest was higher

315 than 0.3. Two areas through which one packet containing a high concentration of pollutants passed showed high probabilities

316 of 0.6 and 0.5. One was southwest of the SMA over the Yellow Sea and the other between North Korea and the coast of

317 northern China over the Yellow Sea.

318

319 One important limitation of the PSCF is its complexity distinguishing between moderate and strong sources. To overcome this

320 problem, we can apply the concentration-weighted trajectory (CWT) method to compute concentration fields for identifying

321 strong source areas of pollutants. The CWT method, based on concentration values over each trajectory, estimates the trajectory

322 weighted concentration in each grid cell by averaging the sample pollutant concentrations of trajectories crossing each grid

323 cell (one degree by one degree). The results of CWT show close agreement with those of the PSCF. Figure 12(d) shows the

324 distribution of weighted trajectory concentrations of CO surrounding the SMA in May 2016. The CWT results show that not

325 only were the Yellow Sea and the Shandong Peninsula potential sources of high concentration over the SMA, but other local

326 sources may also have been strong sources. For example, the Pyongyang area in North Korea had a high concentration,

327 weighted over 250 ppb, indicating a strong potential source of CO in this month. Furthermore, local regions such as those at

328 west, east, and south of the SMA showed a strong potential source of high CO concentrations in Seoul. Among the long-

329 distance sources, only the Shandong Peninsula and some parts of northern China had CO concentrations of around 100 ppmV

330 according to the CWT analysis. As other long-distance sources were not strong sources because of the scarcity of trajectories

331 in these areas, we consider them rare sources. For instance, although the LRT explained the high CO concentrations over the

332 SMA during the extreme pollution period (May 25-28), during longer periods (e.g., one month or one year), with a similar

333 contribution, distant regions from the SMA may not have been strong sources.

334

6   Conclusion

336 In this study, we introduced the C-TRAIL Lagrangian output, extracted from the Eulerian CMAQ model. The comprehensive

337 output of C-TRAIL directly linked the trajectories of pollution from the source to the receptor. We used concentration and

338 trajectory values of C-TRAIL outputs to investigate the pollution status of originated air masses by classifying the outputs

339 from May 2016 over East Asia into separate categories. Unlike the conventional Eulerian CO concentration plots for separate

340 periods, which did not exhibit a clear relationship between the source and the receptor, the C-TRAIL outputs, which combined

341 trajectories and concentrations, more vividly determined the impact of LRT on pollution during the EPP. Furthermore, during

342 the dynamic weather period, C-TRAIL outputs showed that polluted packets from the Shandong Peninsula were responsible

343 for high CO concentrations. The outputs for the SP revealed CO concentrations of less than 100 ppbv for distant packets,

344 strong evidence supporting the link between local trajectories and CO concentrations over the SMA during this period.

More comprehensive investigations on C-TRAIL outputs found that the Shandong Peninsula, local regions near the SMA, and the Pyongyang area were potentially strong sources of CO pollutants during the entire month of May 2016. Overall, by analyzing the trajectory paths of packets that reached specific locations, we were able to generalize that C-TRAIL represents a practical tool for ascertaining the impact of long-range transport on species concentrations over a receptor by simultaneously providing concentrations and trajectories. C-TRAIL can be applied to LRT-impacted regions such as East Asia, North America, and India. Owing to uncertainties inherent in emission inventories and immature diffusion modeling methods, however, C-TRAIL outputs may have limitations that we will address in future work. The objective of this study is to suggest an effective tool for establishing a link between real sources of pollution to a receptor via trajectory analysis. The results of this study over East Asia showed the reliability and various advantages of C-TRAIL output. Therefore, because of its capability to determine the trajectories of masses of CO concentrations, C-TRAIL output could prove to be a highly useful tool for those who model air quality over a specific region and investigate sources of polluted air masses.

Code and Data Availability. The C-TRAIL version 1.0 is derived from the CMAQ v5.2 model. The CMAQ model is an open-source model that can be accessed via Zenodo (https://doi.org/10.5281/zenodo.1167892). The documentation and tutorials on CMAQ are available on the US EPA modeling website https://www.cmascenter.org/ (last access: June 2020). The C-TRAIL v1.0 model source code and pre/post-processing scripts are available via Zenodo (https://doi.org/10.5281/zenodo.3885782). The flight observations used for evaluation of the model were downloaded from the KORUS-AQ website https://www-air.larc.nasa.gov/missions/korus-aq/ (last access: June 2020), DC8 Aircraft observations for CO comparison and Ozonsonde data for wind speed/direction evaluation. For surface concentration evaluations, we used surface observational data from the Air Quality Monitoring Station (AQMS) network operated by the National Institute of Environmental Research (NIER), which is available for the public on https://www.airkorea.or.kr/web (last access: June 2020).

Supplement. A supplementary document related to this article follows.

Author Contribution. A. P., Y. C. and B. S. contributed to the design and implementation of the research. J. J. prepared the CMAQ model and inputs. A. P. prepared the C-TRAIL model, analyzed the results, and took the lead in writing the manuscript. Y. C. and C. H. S. supervised the project. All authors discussed the results and commented on the manuscript and contributed to the final version of the manuscript.

Competing interest. The authors declare no competing financial and/or non-financial interests in relation to the work described.

378   **Acknowledgments.** We wish to acknowledge Dr. Peter Percell for his technical support in the development of CMAQ-TG in

379   this research. This study was supported by the National Strategic Project-Fine particle of the National Research Foundation of

380   Korea (NRF) funded by the Ministry of Science and ICT (MSIT), the Ministry of Environment (ME), and the Ministry of

381   Health and Welfare (MOHW) (NRF-2017M3D8A1092022).

7   References

Al-Saadi, J., Carmichael, G., Crawford, J., Emmons, L., Kim, S., Song, C.-K., Chang, L.-S., Lee, G., Kim, J. and Park, R.: KORUS-AQ: An International Cooperative Air Quality Field Study in Korea (2016). [online] Available from: https://espo.nasa.gov/korus-aq/content/KORUS-AQ, 2016.

Bertschi, I. T. and Jaffe, D. A.: Long-range transport of ozone, carbon monoxide, and aerosols to the NE Pacific troposphere during the summer of 2003: Observations of smoke plumes from Asian boreal fires, J. Geophys. Res. D Atmos., 110(5), 1–14, doi:10.1029/2004JD005135, 2005.

Byun, D. and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system, Appl. Mech. Rev., 59(1–6), 51–76, doi:10.1115/1.2128636, 2006.

Carroll, M., Ocko, I. B., McNeal, F., Weremijewicz, J., Hogg, A. J., Opoku, N., Bertman, S. B., Neil, L., Fortner, E., Thornberry, T., Town, M. S., Yip, G. and Yageman, L.: An Assessment of Forest Pollutant Exposure Using Back Trajectories, Anthropogenic Emissions, and Ambient Ozone and Carbon Monoxide Measurements, Am. Geophys. Union, Fall Meet. 2008, Abstr. id. A41H-0227, 2008.

Carslaw, D. C. and Ropkins, K.: Openair - An r package for air quality data analysis, Environ. Model. Softw., 27–28, 52–61, doi:10.1016/j.envsoft.2011.09.008, 2012.

Chen, T. F., Chang, K. H. and Tsai, C. Y.: Modeling direct and indirect effect of long range transport on atmospheric PM 2.5 levels, Atmos. Environ., 89, 1–9, doi:10.1016/j.atmosenv.2014.01.065, 2014.

Chock, D. P., Sun, P. and Winkler, S. L.: Trajectory-grid: An accurate sign-preserving advection-diffusion approach for air quality modeling, Atmos. Environ., 30(6), 857–868, doi:10.1016/1352-2310(95)00332-0, 1996.

Chock, D. P., Whalen, M. J., Winkler, S. L. and Sun, P.: Implementing the trajectory-grid transport algorithm in an air quality model, Atmos. Environ., 39(22), 4015–4023, doi:10.1016/j.atmosenv.2005.03.037, 2005.

Choi, J., Park, R. J., Lee, H. M., Lee, S., Jo, D. S., Jeong, J. I., Henze, D. K., Woo, J. H., Ban, S. J., Lee, M. Do, Lim, C. S., Park, M. K., Shin, H. J., Cho, S., Peterson, D. and Song, C. K.: Impacts of local vs. trans-boundary emissions from different sectors on PM2.5 exposure in South Korea during the KORUS-AQ campaign, Atmos. Environ., 203, 196–205, doi:10.1016/j.atmosenv.2019.02.008, 2019.

Choi, S. H., Ghim, Y. S., Chang, Y. S. and Jung, K.: Behavior of particulate matter during high concentration episodes in Seoul, Environ. Sci. Pollut. Res., 21(9), 5972–5982, doi:10.1007/s11356-014-2555-y, 2014.

Chuang, M. T., Fu, J. S., Jang, C. J., Chan, C. C., Ni, P. C. and Lee, C. Te: Simulation of long-range transport aerosols from the Asian Continent to Taiwan by a Southward Asian high-pressure system, Sci. Total Environ., 406(1–2), 168–179, doi:10.1016/j.scitotenv.2008.07.003, 2008.

Chuang, M. T., Lee, C. Te and Hsu, H. C.: Quantifying PM2.5 from long-range transport and local pollution in Taiwan during winter monsoon: An efficient estimation method, J. Environ. Manage., 227(July), 10–22, doi:10.1016/j.jenvman.2018.08.066,

415 2018.

416 Cristofanelli, P., Bonasoni, P., Carboni, G., Calzolari, F., Casarola, L., Zauli Sajani, S. and Santaguida, R.: Anomalous high

417 ozone concentrations recorded at a high mountain station in Italy in summer 2003, Atmos. Environ., 41(7), 1383–1394,

418 doi:10.1016/j.atmosenv.2006.10.017, 2007.

419 Döös, K., Jönsson, B. and Kjellsson, J.: Evaluation of oceanic and atmospheric trajectory schemes in the TRACMASS

420 trajectory model v6.0, Geosci. Model Dev., 10(4), 1733–1749, doi:10.5194/gmd-10-1733-2017, 2017.

421 Draxler, R. R.: An overview of the HYSPLIT_4 modelling system for trajectories, dispersion and deposition, Aust. Meteorol.

422 Mag., 47(4), 295–308, 1998.

423 Eslami, E., Salman, A. K., Choi, Y., Sayeed, A. and Lops, Y.: A data ensemble approach for real-time air quality forecasting

424 using extremely randomized trees and deep neural networks, Neural Comput. Appl., doi:10.1007/s00521-019-04287-6, 2019.

425 Gratz, L. E., Jaffe, D. A. and Hee, J. R.: Causes of increasing ozone and decreasing carbon monoxide in springtime at the Mt.

426 Bachelor Observatory from 2004 to 2013, Atmos. Environ., 109, 323–330, doi:10.1016/j.atmosenv.2014.05.076, 2015.

427 Halliday, H. S., DiGangi, J. P., Choi, Y., Diskin, G. S., Pusede, S. E., Rana, M., Nowak, J. B., Knote, C., Ren, X., He, H.,

428 Dickerson, R. R. and Li, Z.: Using Short-Term CO/CO 2 Ratios to Assess Air Mass Differences over the Korean Peninsula

429 during KORUS-AQ , J. Geophys. Res. Atmos., 1–22, doi:10.1029/2018jd029697, 2019.

430 Heald, C. C., Jacob, D. J., Fiore, A. M., Emmons, L. K., Gille, J. C., Deeter, M. N., Warner, J., Edwards, D. P., Crawford, J.

431 H., Hamlin, A. J., Sachse, G. W., Browell, E. V., Avery, M. A., Vay, S. A., Westberg, D. J., Blake, D. R., Singh, H. B.,

432 Sandholm, S. T., Talbot, R. W. and Fuelberg, H. E.: Asian outflow and trans-Pacific transport of carbon monoxide and ozone

433 pollution: An integrated satellite, aircraft, and model perspective, J. Geophys. Res. D Atmos., 108(24),

434 doi:10.1029/2003jd003507, 2003.

435 Hu, Y. and Talat Odman, M.: A comparison of mass conservation methods for air quality models, Atmos. Environ., 42(35),

436 8322–8330, doi:10.1016/j.atmosenv.2008.07.042, 2008.

437 Jeon, W., Choi, Y., Percell, P., Hossein Souri, A., Song, C. K., Kim, S. T. and Kim, J.: Computationally efficient air quality

438 forecasting tool: Implementation of STOPS v1.5 model into CMAQ v5.0.2 for a prediction of Asian dust, Geosci. Model Dev.,

439 9(10), 3671–3684, doi:10.5194/gmd-9-3671-2016, 2016.

440 Jung, J., Souri, A. H., Wong, D. C., Lee, S., Jeon, W., Kim, J. and Choi, Y.: The Impact of the Direct Effect of Aerosols on

441 Meteorology and Air Quality Using Aerosol Optical Depth Assimilation During the KORUS-AQ Campaign, J. Geophys. Res.

442 Atmos., 124(14), 8303–8319, doi:10.1029/2019jd030641, 2019.

443 Kain, J. S.: The Kain - Fritsch convective parameterization: An update, J. Appl. Meteorol., 43(1), 170–181, doi:10.1175/1520-

444 0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

445 Kruse, S., Gerdes, A., Kath, N. J. and Herzschuh, U.: Implementing spatially explicit wind-driven seed and pollen dispersal in

446 the individual-based larch simulation model: LAVESI-WIND 1.0, Geosci. Model Dev., 11(11), 4451–4467, doi:10.5194/gmd-

447 11-4451-2018, 2018.

448 Lee, S., Ho, C. H. and Choi, Y. S.: High-PM10 concentration episodes in Seoul, Korea: Background sources and related

449   meteorological conditions, Atmos. Environ., 45(39), 7240–7247, doi:10.1016/j.atmosenv.2011.08.071, 2011.

450   Lee, S., Ho, C. H., Lee, Y. G., Choi, H. J. and Song, C. K.: Influence of transboundary air pollutants from China on the high-
451   PM10 episode in Seoul, Korea for the period October 16-20, 2008, Atmos. Environ., 77, 430–439,
452   doi:10.1016/j.atmosenv.2013.05.006, 2013.

453   Lee, S., Kim, J., Choi, M., Hong, J., Lim, H., Eck, T. F., Holben, B. N., Ahn, J. Y., Kim, J. and Koo, J. H.: Analysis of long-
454   range transboundary transport (LRTT) effect on Korean aerosol pollution during the KORUS-AQ campaign, Atmos. Environ.,
455   204(February), 53–67, doi:10.1016/j.atmosenv.2019.02.020, 2019.

456   Li, M., Zhang, Q., Kurokawa, J. I., Woo, J. H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng,
457   Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H. and Zheng, B.: MIX: A mosaic Asian anthropogenic emission
458   inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17(2), 935–
459   963, doi:10.5194/acp-17-935-2017, 2017.

460   Liu, Y., Xu, S., Ling, T., Xu, L. and Shen, W.: Heme oxygenase/carbon monoxide system participates in regulating wheat
461   seed germination under osmotic stress involving the nitric oxide pathway, J. Plant Physiol., 167(16), 1371–1379,
462   doi:10.1016/j.jplph.2010.05.021, 2010.

463   Lops, Y., Choi, Y., Eslami, E. and Sayeed, A.: Real-time 7-day forecast of pollen counts using a deep convolutional neural
464   network, Neural Comput. Appl., 1–10, doi:10.1007/s00521-019-04665-0, 2019.

465   Miyazaki, K., Sekiya, T., Fu, D., Bowman, K. W., Kulawik, S. S., Sudo, K., Walker, T., Kanaya, Y., Takigawa, M., Ogochi,
466   K., Eskes, H., Boersma, K. F., Thompson, A. M., Gaubert, B., Barre, J. and Emmons, L. K.: Balance of Emission and
467   Dynamical Controls on Ozone During the Korea-United States Air Quality Campaign From Multiconstituent Satellite Data
468   Assimilation, J. Geophys. Res. Atmos., 124(1), 387–413, doi:10.1029/2018JD028912, 2019.

469   Oh, H. R., Ho, C. H., Kim, J., Chen, D., Lee, S., Choi, Y. S., Chang, L. S. and Song, C. K.: Long-range transport of air
470   pollutants originating in China: A possible major cause of multi-day high-PM10 episodes during cold season in Seoul, Korea,
471   Atmos. Environ., 109, 23–30, doi:10.1016/j.atmosenv.2015.03.005, 2015.

472   Pekney, N. J., Davidson, C. I., Zhou, L. and Hopke, P. K.: Application of PSCF and CPF to PMF-Modeled Sources of PM $_{2.5}$
473   in Pittsburgh, Aerosol Sci. Technol., 40(10), 952–961, doi:10.1080/02786820500543324, 2006.

474   Petetin, H., Beekmann, M., Sciare, J., Bressi, M., Rosso, A., Sanchez, O. and Ghersi, V.: A novel model evaluation approach
475   focusing on local and advected contributions to urban PM2.5 levels - Application to Paris, France, Geosci. Model Dev., 7(4),
476   1483–1505, doi:10.5194/gmd-7-1483-2014, 2014.

477   Price, H. U., Jaffe, D. A., Cooper, O. R. and Doskey, P. V.: Photochemistry, ozone production, and dilution during long-range
478   transport episodes from Eurasia to the northwest United States, J. Geophys. Res. D Atmos., 109(23), 1–10,
479   doi:10.1029/2003JD004400, 2004.

480   Pu, W., Zhao, X., Shi, X., Ma, Z., Zhang, X. and Yu, B.: Impact of long-range transport on aerosol properties at a regional
481   background station in Northern China, Atmos. Res., 153, 489–499, doi:10.1016/j.atmosres.2014.10.010, 2015.

482   Rößler, T., Stein, O., Heng, Y., Baumeister, P. and Hoffmann, L.: Trajectory errors of different numerical integration schemes

diagnosed with the MPTRAC advection module driven by ECMWF operational analyses, Geosci. Model Dev., 11(2), 575–592, doi:10.5194/gmd-11-575-2018, 2018.

Sadeghi, B., Choi, Y., Yoon, S., Flynn, J., Kotsakis, A. and Lee, S.: The characterization of fine particulate matter downwind of Houston: Using integrated factor analysis to identify anthropogenic and natural sources, Environ. Pollut., 262, 114345, doi:10.1016/j.envpol.2020.114345, 2020.

Salvador, P., Artíñano, B., Querol, X. and Alastuey, A.: A combined analysis of backward trajectories and aerosol chemistry to characterise long-range transport episodes of particulate matter: The Madrid air basin, a case study, Sci. Total Environ., 390(2–3), 495–506, doi:10.1016/j.scitotenv.2007.10.052, 2008.

Sarwar, G., Simon, H., Bhave, P. and Yarwood, G.: Examining the impact of heterogeneous nitryl chloride production on air quality across the United States, Atmos. Chem. Phys., 12(14), 6455–6473, doi:10.5194/acp-12-6455-2012, 2012.

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A. and Jung, J.: Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance, Neural Networks, 121, 396–408, doi:10.1016/j.neunet.2019.09.033, 2020.

Souri, A. H., Choi, Y., Li, X., Kotsakis, A. and Jiang, X.: A 15-year climatology of wind pattern impacts on surface ozone in Houston, Texas, Atmos. Res., 174–175, 124–134, doi:10.1016/j.atmosres.2016.02.007, 2016.

Stenke, A., Dameris, M., Grewe, V. and Garny, H.: Implications of lagrangian transport for simulations with a coupled chemistry-climate model, Atmos. Chem. Phys., 9(15), 5489–5504, doi:10.5194/acp-9-5489-2009, 2009.

Stohl, A.: Trajectory statistics - A new method to establish source-receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe, Atmos. Environ., 30(4), 579–587, doi:10.1016/1352-2310(95)00314-2, 1996.

Stohl, A.: Computation, accuracy and applications of trajectories- a review and bibliography, Dev. Environ. Sci., 1(C), 615–654, doi:10.1016/S1474-8177(02)80024-9, 2002.

Stohl, A. and Seibert, P.: Accuracy of trajectories as determined from the conservation of meteorological tracers, Q. J. R. Meteorol. Soc., 124(549), 1465–1484, doi:10.1002/qj.49712454907, 1998.

Vay, S. A., Choi, Y., Vadrevu, K. P., Blake, D. R., Tyler, S. C., Wisthaler, A., Hecobian, A., Kondo, Y., Diskin, G. S., Sachse, G. W., Woo, J. H., Weinheimer, A. J., Burkhart, J. F., Stohl, A. and Wennberg, P. O.: Patterns of CO_2 and radiocarbon across high northern latitudes during International Polar Year 2008, J. Geophys. Res. Atmos., 116(14), 1–22, doi:10.1029/2011JD015643, 2011.

Wang, F., Chen, D. S., Cheng, S. Y., Li, J. B., Li, M. J. and Ren, Z. H.: Identification of regional atmospheric PM10 transport pathways using HYSPLIT, MM5-CMAQ and synoptic pressure pattern analysis, Environ. Model. Softw., 25(8), 927–934, doi:10.1016/j.envsoft.2010.02.004, 2010.

Weiss-Penzias, P., Jaffe, D. A., Jaeglé, L. and Liang, Q.: Influence of long-range-transported pollution on the annual and diurnal cycles of carbon monoxide and ozone at Cheeka Peak Observatory, J. Geophys. Res. D Atmos., 109(23), 1–15, doi:10.1029/2004JD004505, 2004.

Xu, S., Warner, N., Bohlin-Nizzetto, P., Durham, J. and McNett, D.: Long-range transport potential and atmospheric persistence of cyclic volatile methylsiloxanes based on global measurements, Chemosphere, 228, 460–468,

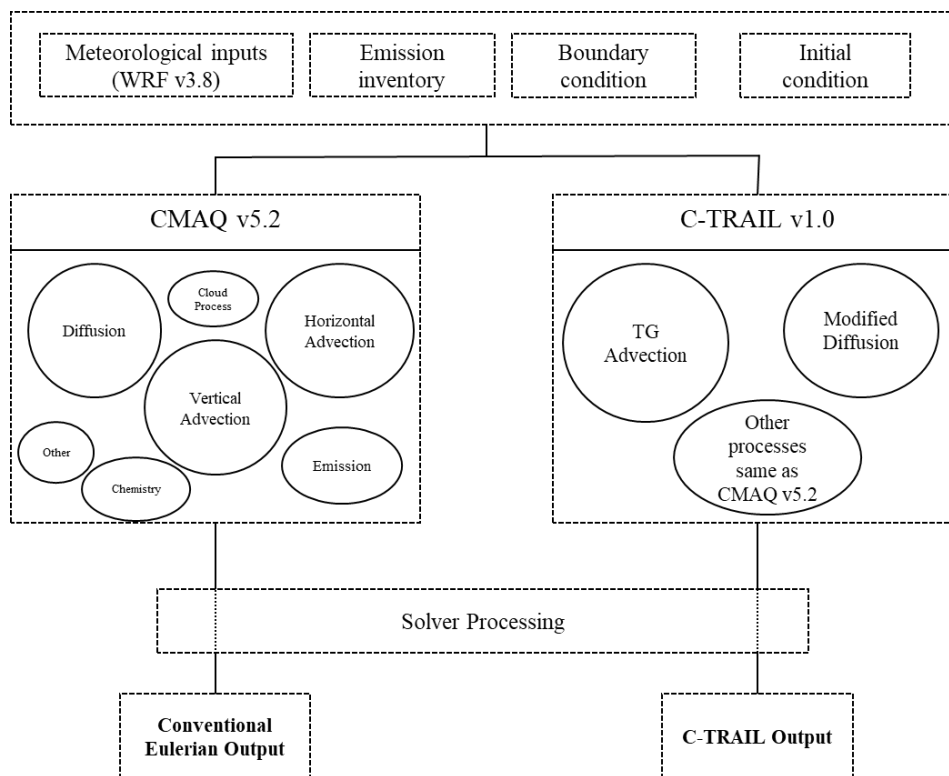
517   doi:10.1016/j.chemosphere.2019.04.130, 2019.

518   Zhang, Q., Xue, D., Liu, X., Gong, X. and Gao, H.: Process analysis of PM 2.5 pollution events in a coastal city of China using

519   CMAQ, J. Environ. Sci. (China), 79, 225–238, doi:10.1016/j.jes.2018.09.007, 2019.

520

521

522

523
524 **Figure 1: Schematic of conventional CMAQ versus C-TRAIL**

525

1. **Initialization**

    - Packets are generated based on the IC and BC in grid cells.

    - Each packet receives an ID, x, y, z.

    - Time-step selection is based on the condition that a packet may not travel more than ¾ of the distance between all opposite faces of a grid cell.

2. **Packet Management**

    - The number of packets in each grid cell must not exceed five.

        If exceeded: Remove the extra packets.

        The concentration of remaining packets will be the average of all available packets.

    - Each grid cell must not be empty.

        If empty: Add an extra packet.

        The concentration of the added packet will be similar to that of the closest packet.

    - The properties of the packets are controlled through all time steps.

3. **Advection**

    - The three-dimensional advection equation is solved to update the location of packets.

4. **Diffusion**

    - The implicit Eulerian diffusion equation is solved to obtain an average over the number of packets in each grid cell.

    - By considering each packet the center of the cell and the cell average as neighboring cells, the diffusion equation is solved for each packet using the predictor-corrector method.

    - Horizontal diffusion requires extra sub-grid diffusion for pair-wise diffusion.

5. **Emission**

    - The emission fluxes of various species (similar to CMAQ) are added to each packet through vertical diffusion.

6. **Output Generation**

    - Based on Lambert's projection, the x, y, and z of each packet are converted into longitude, latitude, and altitude.

    - C-TRAIL is generated during every output time step (1 hour).

527   **Figure 2: Algorithm of the C-TRAIL model**

528

529

530

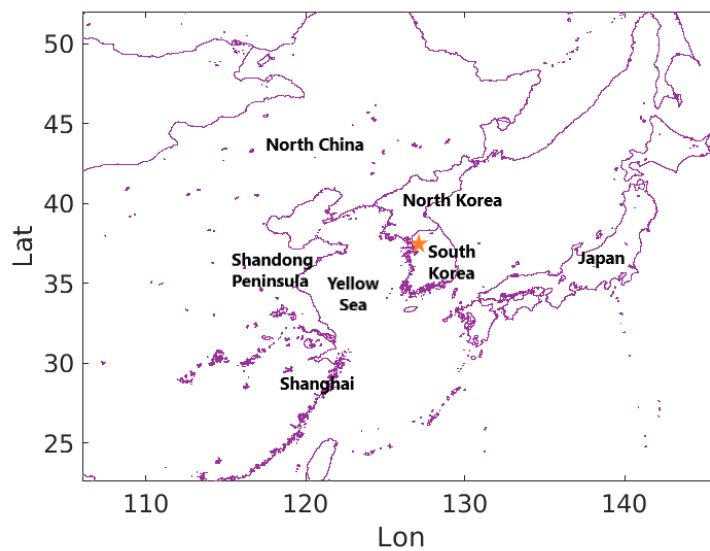



Figure 3: Domain of the study; the orange star indicates the Seoul Metropolitan Area (SMA)

532

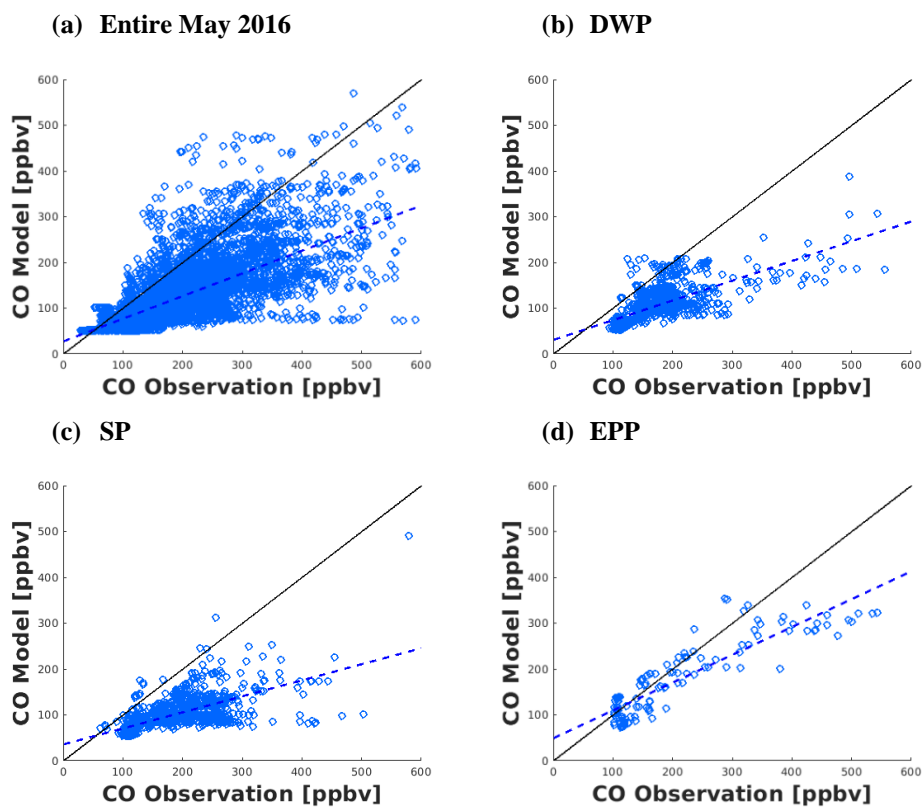
533

Table 1: Comparison of the statistical parameters of CMAQ CO concentrations to aircraft measurements
(COR: correlation, IOA: index of agreement, RMSE: root mean square error, MAE: mean absolute error)

		Abbreviation	COR	IOA	RMSE	MAE
a)	Entire month of May 2016		0.71	0.72	91.3	66.7
b)	Dynamic Weather Period	DWP	0.72	0.62	81.5	66.2
c)	Stagnant Period	SP	0.65	0.58	98.4	83.3
d)	Extreme Pollution Period	EPP	0.89	0.88	68.7	47.7

536

537

538

539   **Figure 4: CMAQ model results versus aircraft CO measurements for (a) the entire month of May 2016 (n=6865),**
540   **(b) the DWP (n=1750), (c) the SP (n=1548), and (d) the EPP (n=264)**

541

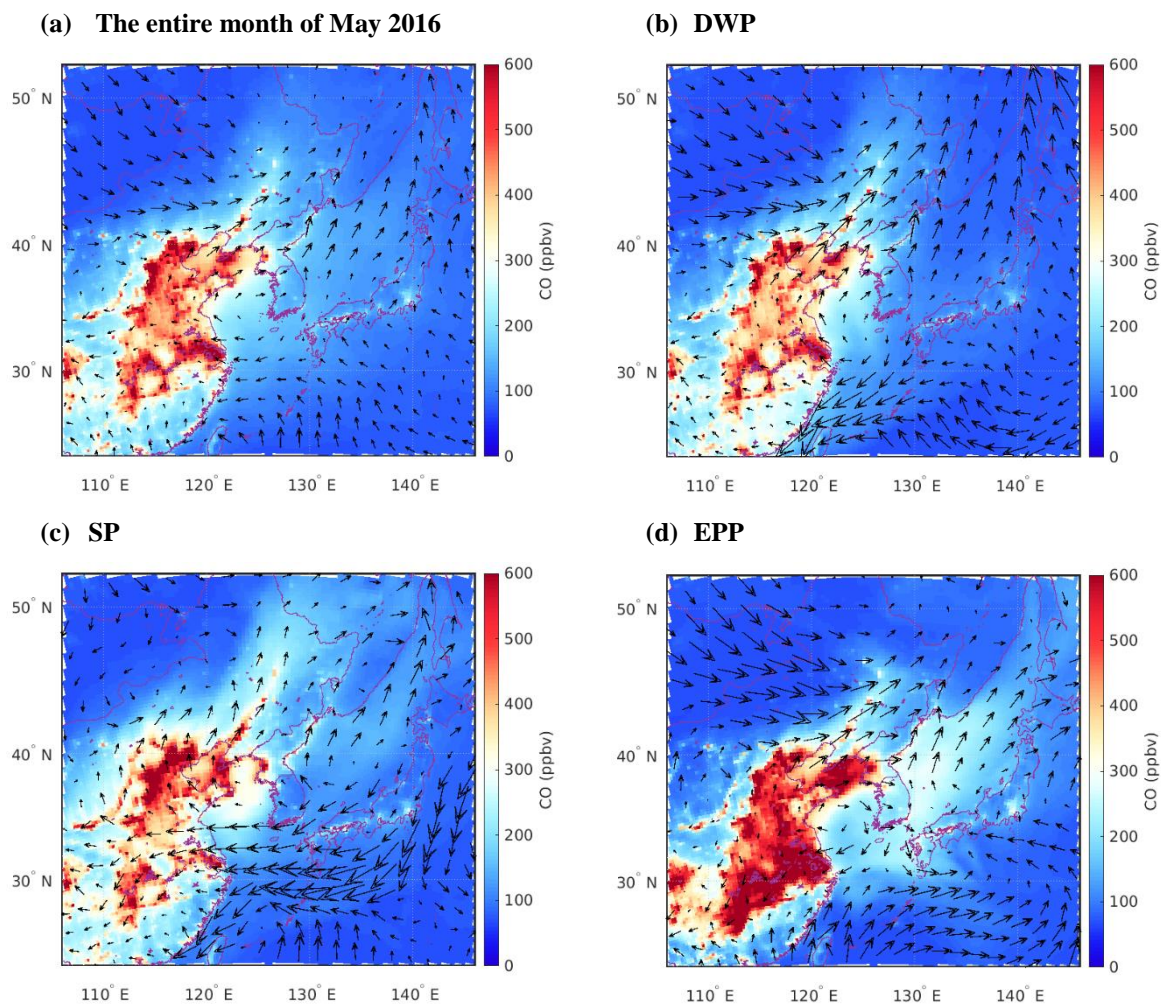



Figure 5: Model CO concentrations and wind patterns over the surface during (a) the entire month of May 2016, (b) the DWP, (c) the SP, and (d) the EPP

546

(a)

(b)

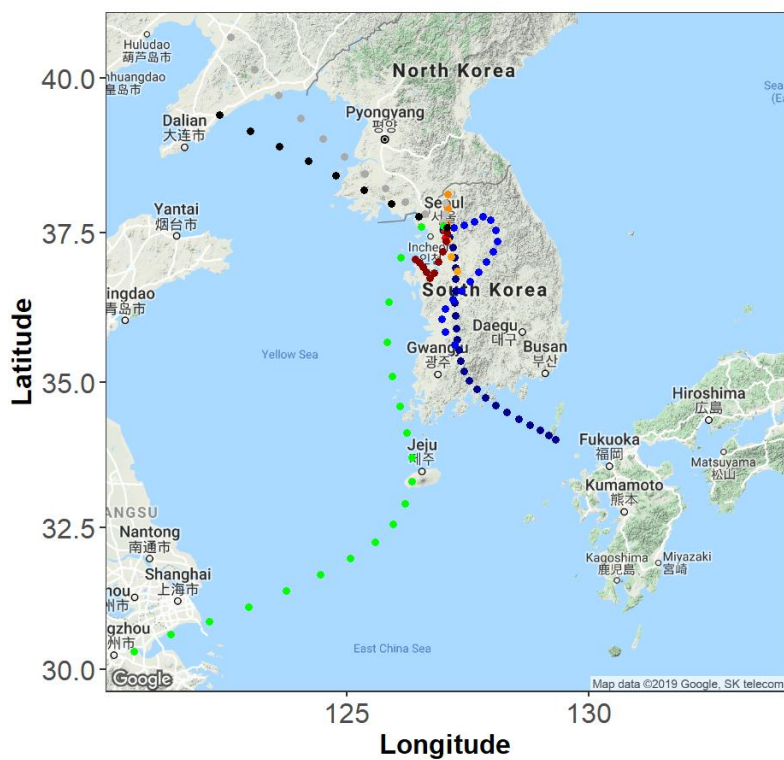
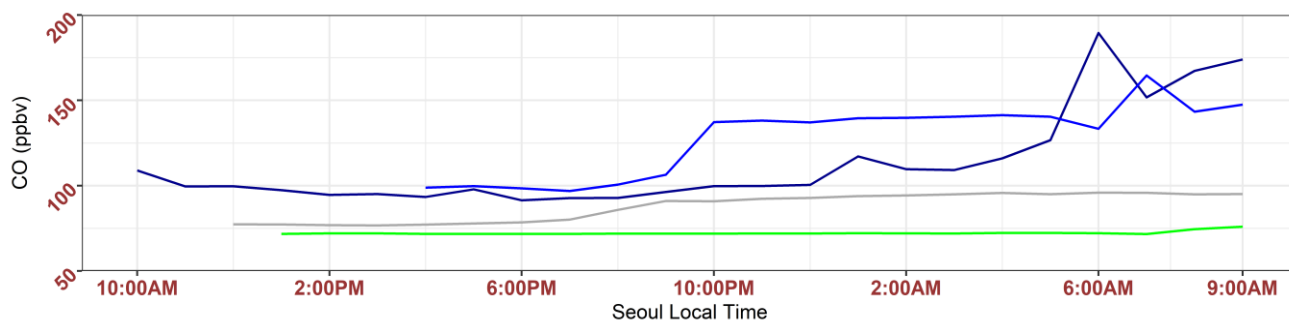



Figure 6: C-TRAIL output for June 4, 2016: (a) the trajectory of packets reaching Seoul at 9:00 AM local time
(b) changes in the CO concentrations of four aged packets moving toward Seoul from source points

549

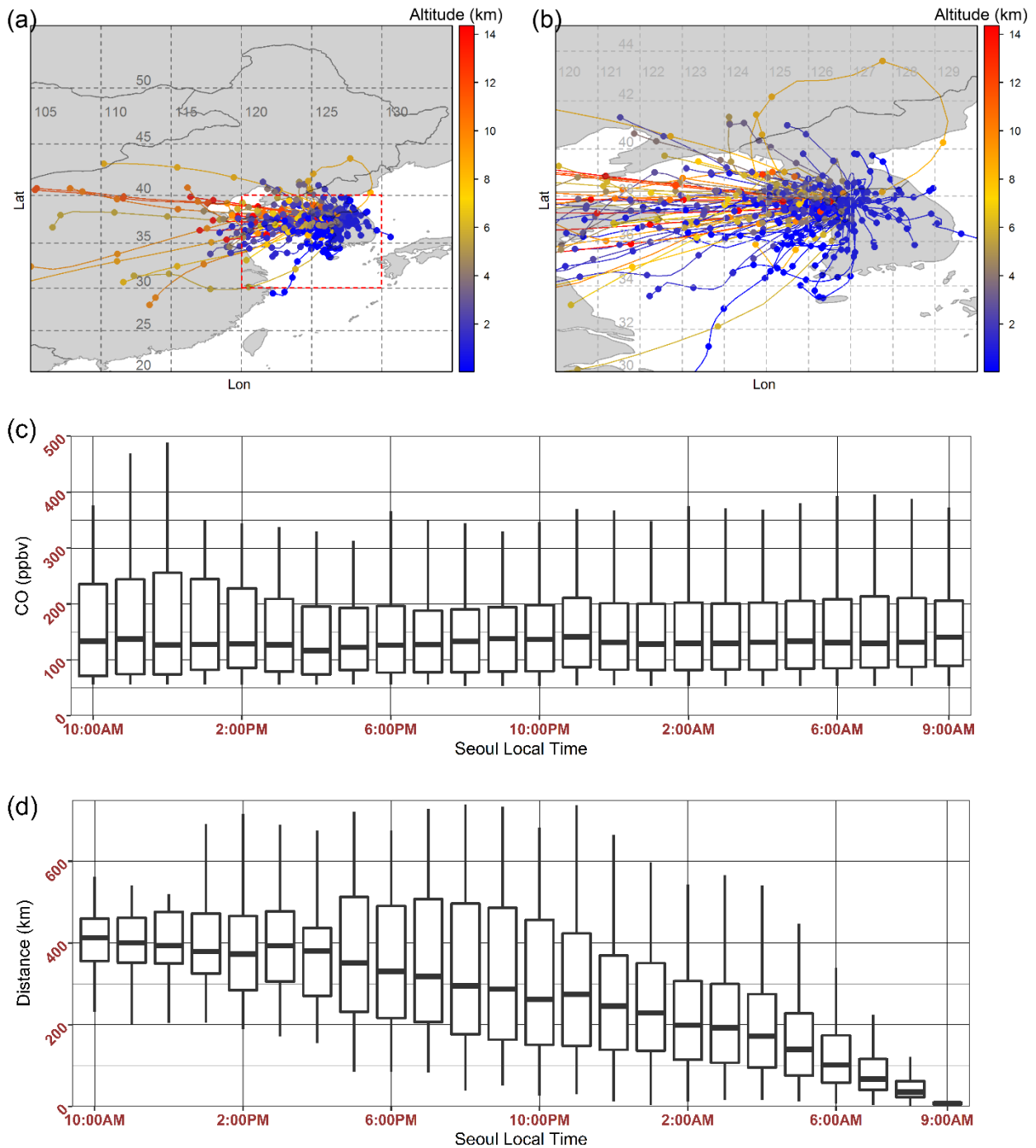



Figure 7: C-TRAIL output for the entire month of May 2016 for Seoul as the receptor: (a) 24-hour trajectories of packets for the entire domain, (b) 24-hour trajectories of packets for the zoomed area in South Korea, (c) the boxplot of the CO concentrations of all packets at each hour before they reached Seoul, and (d) the boxplot of packet distances from Seoul at each hour before the packets reached Seoul

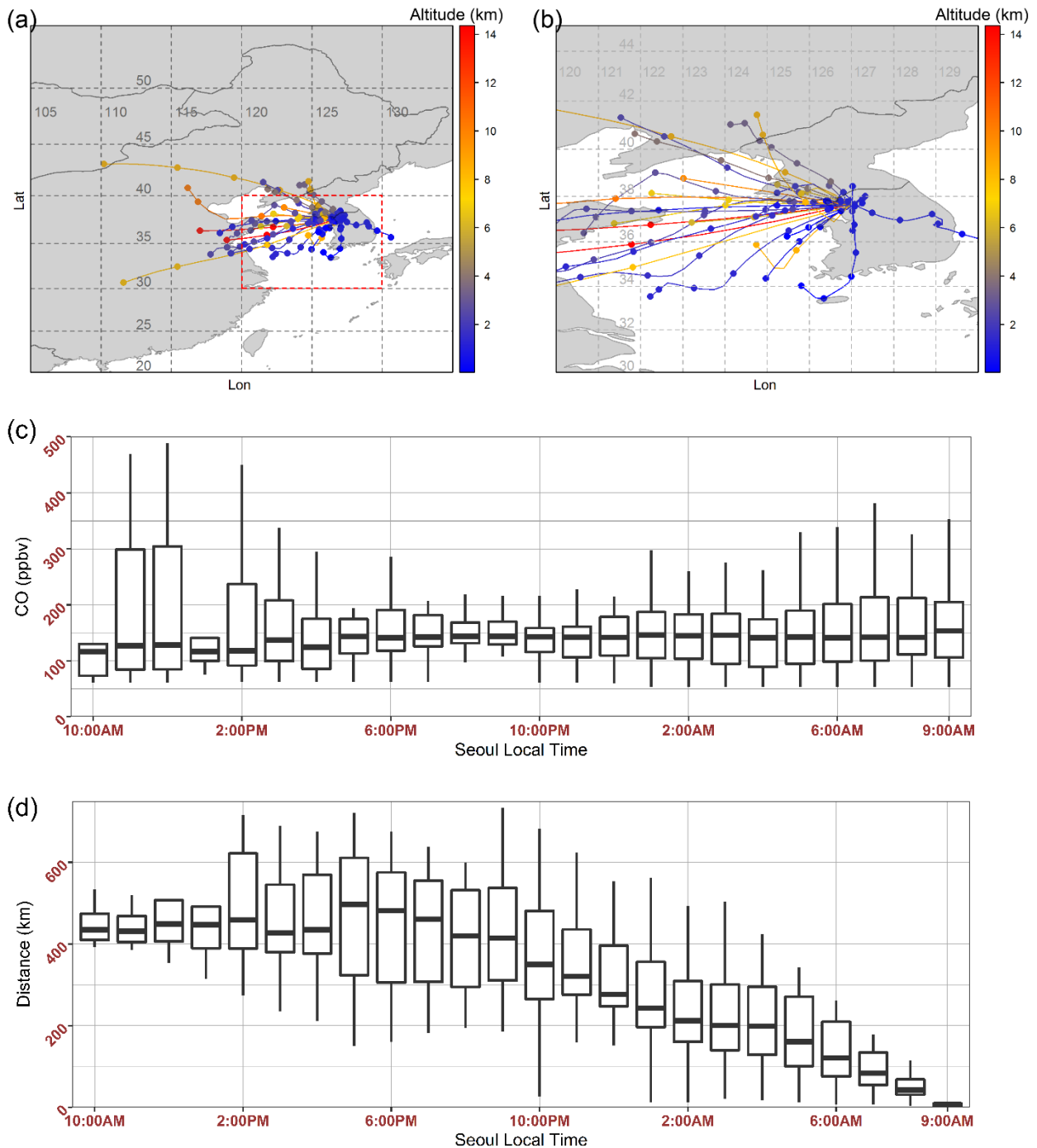
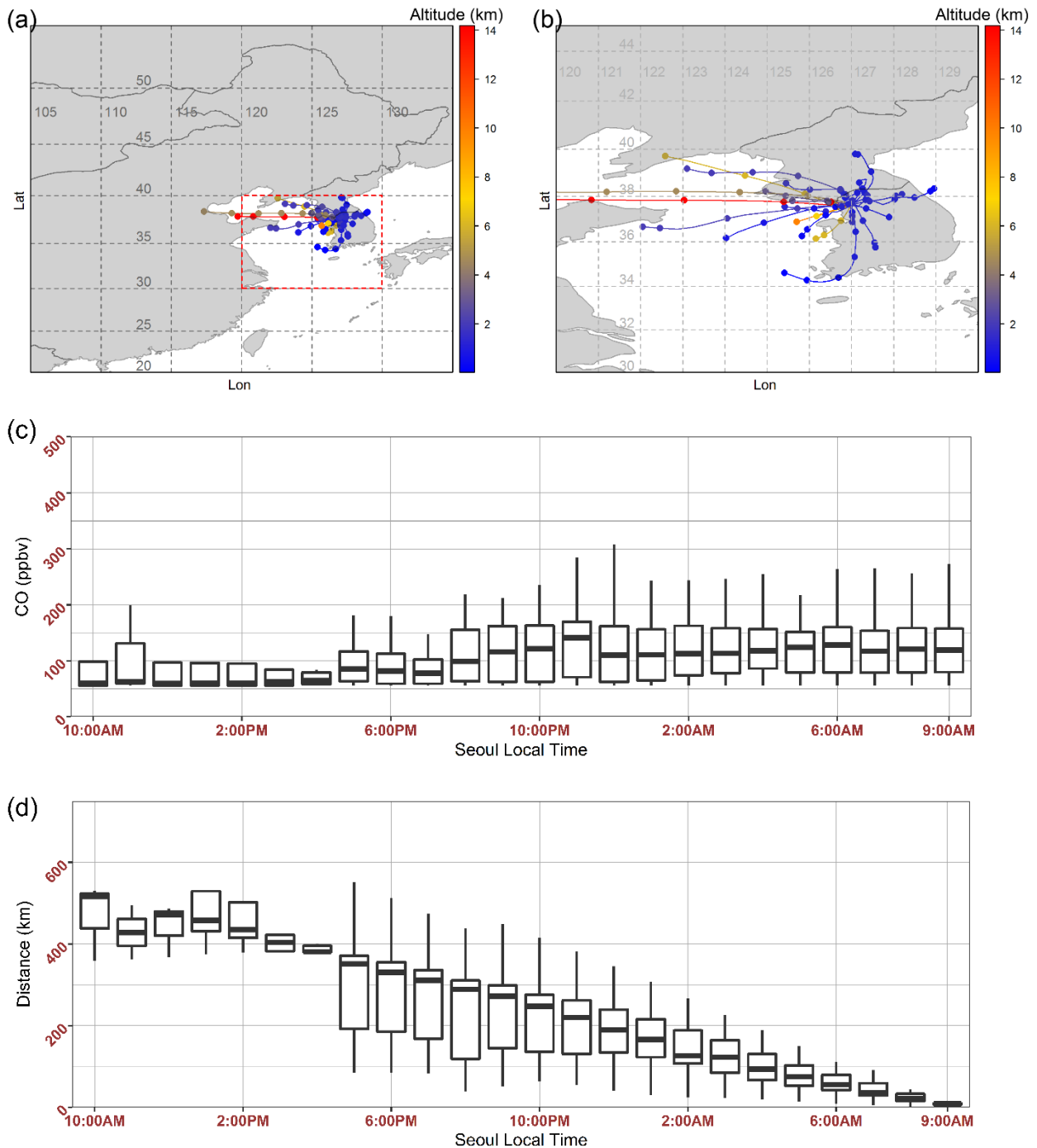



Figure 8: C-TRAIL output for the dynamic weather period (DWP) for Seoul as the receptor: (a) 24-hour trajectories of packets for the entire domain, (b) 24-hour trajectories of packets for the zoomed area in South Korea, (c) the boxplot of the CO concentrations of all packets at each hour before they reached Seoul, and (d) the boxplot of packet distances from Seoul at each hour before the packets reached Seoul

558 **Figure 9: C-TRAIL output for the stagnant period (SP) for Seoul as the receptor: (a) 24-hour trajectories of packets for the entire**
559 **domain, (b) 24-hour trajectories of packets for the zoomed area in South Korea, (c) the boxplot of the CO concentrations of all**
560 **packets at each hour before they reached Seoul, and (d) the boxplot of packet distances from Seoul at each hour before the packets**
561 **reached Seoul**

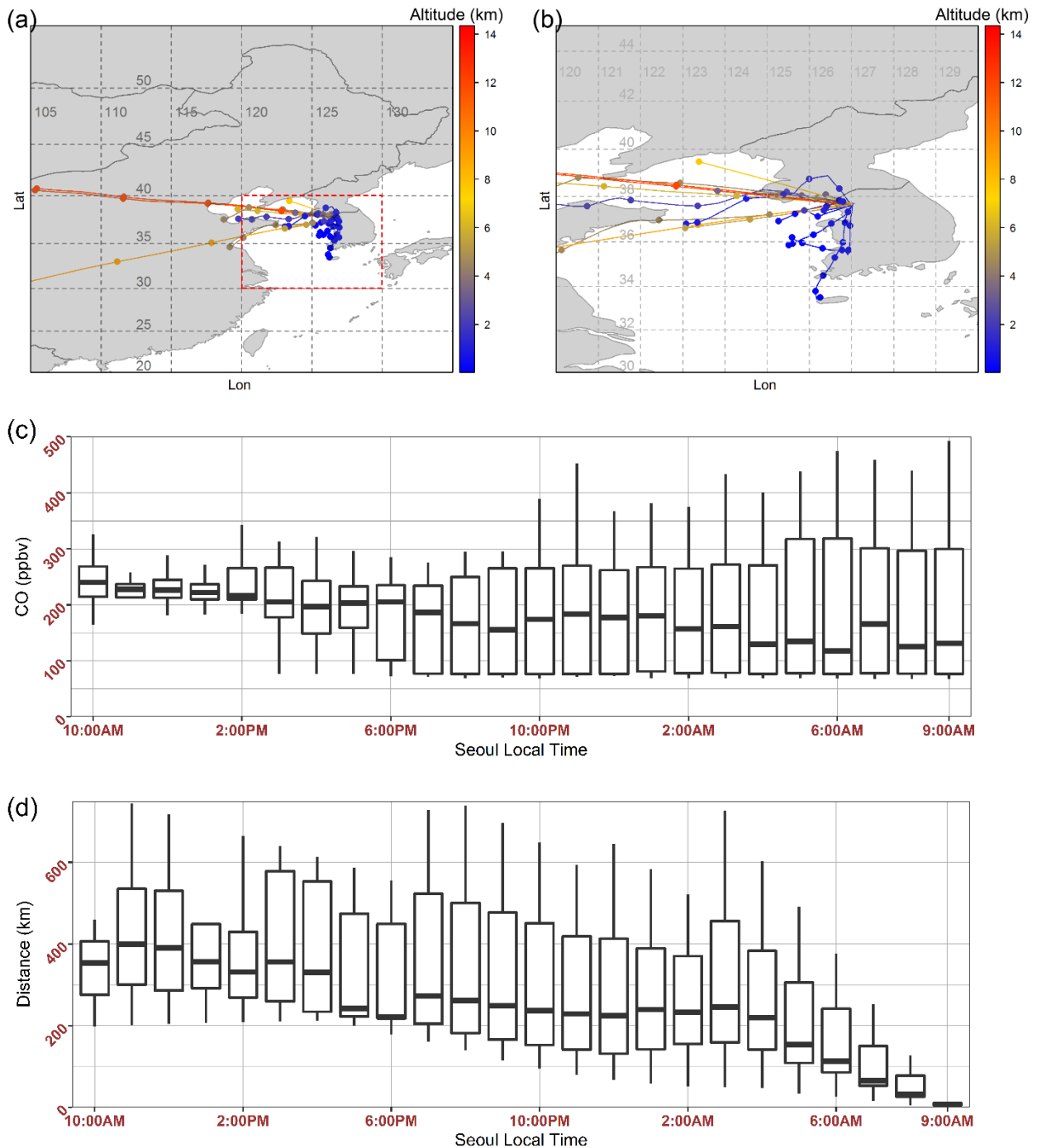
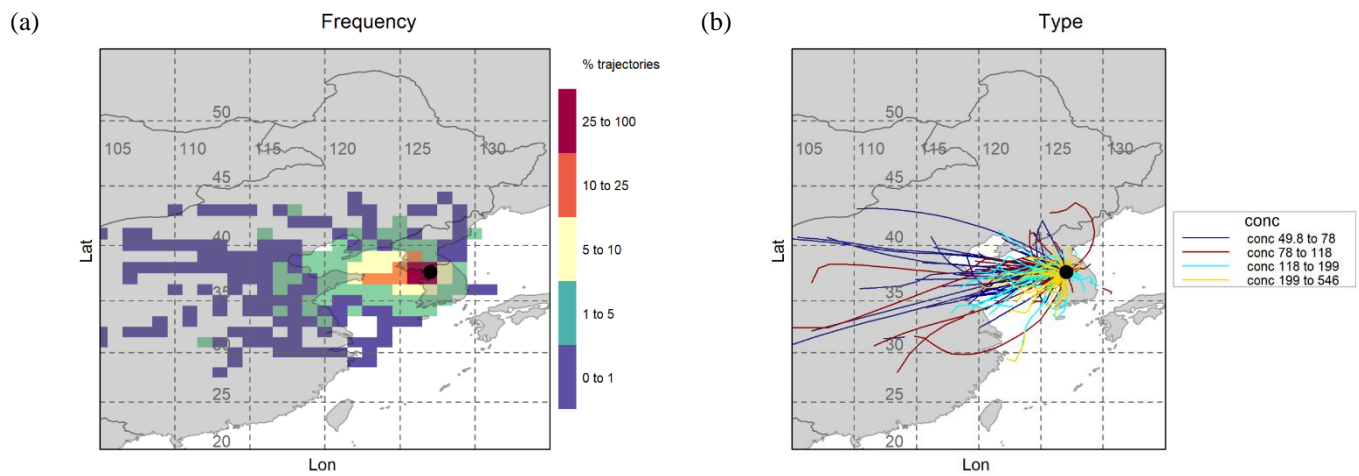



Figure 10: C-TRAIL output for the extreme pollution period (EPP) for Seoul as the receptor: (a) 24-hour trajectories of packets for the entire domain, (b) 24-hour trajectories of packets for the zoomed area in South Korea, (c) the boxplot of the CO concentrations of all packets at each hour before they reached Seoul, and (d) the boxplot of packet distances from Seoul at each hour before the packets reached Seoul

566   **Figure 11: (a) Plot of the frequency of trajectories and (b) the trajectories, classified by their concentration values**

567

568

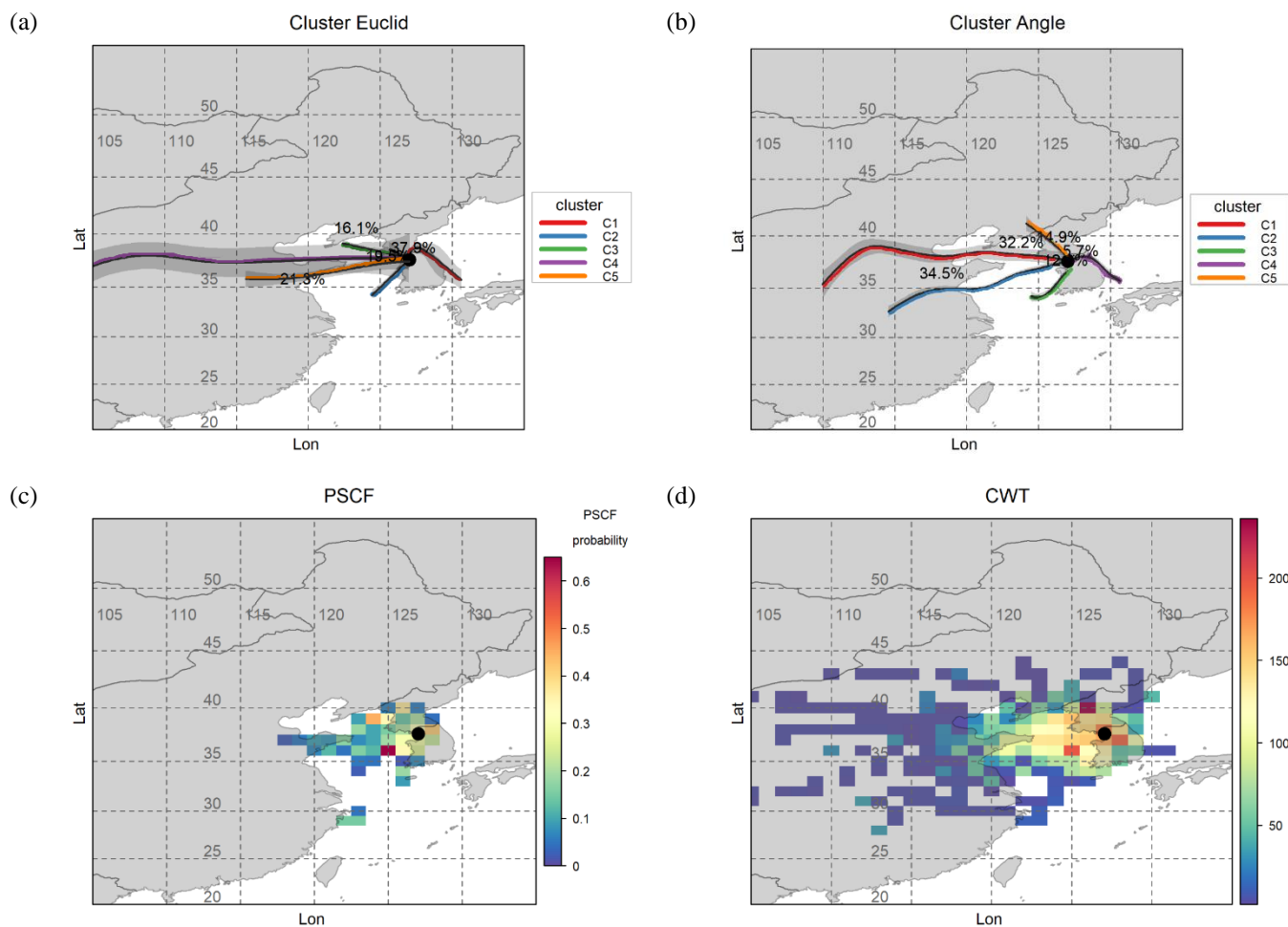



Figure 12: (a) Trajectories clustered by the Euclidian distance function, (b) trajectories clustered by the angle distance function, (c) the potential source contribution factor plot, and (d) the concentration-weighted trajectory plot

569
570

571

572