Performance of Back-Trajectory Statistical Methods and Inverse Modeling Method in Locating Emission Sources

Performance of back-trajectory statistical methods and inverse modeling method in locating emission sources

Xuekun Fang[*,†,‡], Takuya Saito[§], Sunyoung Park[∇], Shanlan Li[∇], Yoko Yokouchi[§], Ronald G. Prinn[†].

[†]Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[‡]College of Environmental Sciences and Engineering, Peking University, Beijing, China

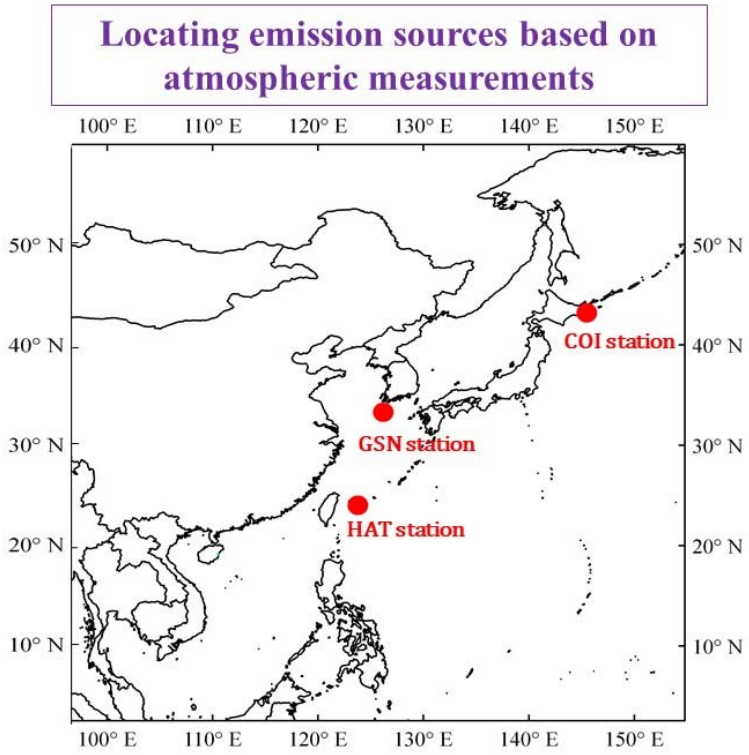
[§]National Institute for Environmental Studies, Tsukuba, Japan

[∇]Department of Oceanography, Kyungpook National University, Daegu, South Korea

Keywords:

Trajectory statistical methods; Inverse modeling; Source attribution; Emission source; FLEXPART; HYSPLIT; Trajectory

16 **TOC**


Locating emission sources based on atmospheric measurements
100° E
110° E
120° E
130° E
140° E
150° E
50° N
40° N
30° N
20° N
10° N
COI station
GSN station
HAT station

17

ABSTRACT

Back-trajectory statistical methods, e.g., potential source contribution functions (PSCF) and concentration-weighted trajectory (CWT) methods, have been widely used in previous studies to locate emission source regions of air pollutants or greenhouse gases. Inverse modeling methods have been developed and used in an increasing number of applications. To this date, there are no comparisons of performance between back-trajectory statistical and inverse modeling methods. This study evaluates the performance of PSCF, CWT and inverse modeling methods by taking advantage of precisely-known locations of trifluoromethane (CHF_3; HFC-23) sources. Results show poor performance of the PSCF and CWT methods and good performance of the inverse modeling method. This study suggests that in studies with the purpose of locating emission source regions, the PSCF and CWT methods should be applied with caution in future studies, and that the inverse modeling method is encouraged to be used much more widely.

INTRODUCTION

Back-trajectories in combination with air quality measurements have been widely used to identify potential emission source areas of atmospheric trace substances (e.g., gases and particulate matter) and to determine respective contributions at measurement sites.[1 and references therein] Various statistical methods based on back-trajectories have been developed. One such method is the potential source contribution function (PSCF) method which is based on the evaluation, for each grid cell (or area), of the ratio of the total number of marked trajectory segment endpoints to the total number of trajectory segment endpoints over that grid cell (see details in Methods section). The PSCF method and its variations have been applied widely in the last three decades, e.g., for aerosol sources in Arctic,[2] sulfur sources in western U.S.A.,[3] hydrofluorocarbon sources in Europe,[4] and black carbon sources in southern China.[5] A second method is the concentration-weighted trajectory (CWT) method, in which each grid cell is assigned a weighted concentration (see details in Methods section). The CWT method has also been widely applied, e.g., for polychlorinated biphenyl sources in Chicago, U.S.A.,[6] halogenated greenhouse gases sources in Europe,[7] and PM_{10}, SO_2 and NO_2 sources in western China.[8]

Inverse modeling is another method that can be used to locate emission source regions and to determine respective emissions strengths. Inverse modeling method does not use back-trajectories but instead uses source-receptor-relationships (SRR; or called emission sensitivity) in combination with atmospheric mole fraction

measurements. The SRR matrix is derived from atmospheric transport models, e.g., FLEXible PARTicle dispersion (FLEXPART) model[9] and The Met Office's Numerical Atmospheric-dispersion Modelling Environment (NAME) model[10]. For example, FLEXPART-based inverse modeling has been used to study hydrochlorofluorocarbon and hydrofluorocarbon sources across the globe,[11] methane in California, U.S.A,[12] methyl chloroform in Europe,[13] and carbon dioxide in East Asia.[14]

Numerous studies have employed PSCF and CWT methods.[1 and references therein] Some of these studies have cross-checked the inferred emission source area field with bottom-up emission information, e.g., the European Monitoring and Evaluation Programme (EMEP) SO_2 inventory[15], mercury emissions inventories for North America[16], smoke from 1998 biomass burning in Central America[17], and the notable 2002 Quebec forest fire[18]. Some studies comparing PSCF and CWT methods have been reported, e.g., by Kabashnikov et al.[19]; however, there are no studies reporting comparisons between PSCF/CWT methods and inverse modeling method with well-known bottom-up emissions information. This study takes advantage of precisely-known locations of trifluoromethane (CHF_3; HFC-23) sources to evaluate the performance of these three methods in locating the source regions. HFC-23 is a potent greenhouse gas with an atmospheric lifetime of ~270 years and a 100-year global warming potential (GWP) of 14800.[20] Thus, HFC-23 is regulated under the Kyoto Protocol and recently became regulated under the Montreal Protocol. HFC-23 is an

74 unavoidable byproduct of the production of chlorodifluoromethane ($CHClF_2$; HCFC-
75 22). Significant amounts of HFC-23 were emitted in East Asia, mainly in China.[21]
76 HFC-23 has negligible direct use, and therefore in East Asia, HFC-23 is mostly
77 emitted as a waste gas at dozens of HCFC-22 production factories.

78 This study evaluates the performance of PSCF, CWT and inverse modeling
79 methods in locating HFC-23 sources, using *in situ* high frequency measurements of
80 atmospheric HFC-23 mole fractions in three East Asian stations and the precisely-
81 known HFC-23 source locations in East Asia.

82 # METHODS

83 ## Atmospheric Measurement

84 This study used HFC-23 measurement data from three stations (Figure 1a): Gosan
85 (GSN; 33.28°N, 126.17°E; air intake height of 17 m above ground level (agl)),
86 Hateruma (HAT; 24.06°N, 123.81°E; air intake height of 36.5 m agl) and Cape Ochi-
87 ishi (COI; 43.16°N, 145.50°E; air intake height of 51 m agl). These three stations are
88 a part of the Advanced Global Atmospheric Gases Experiment network (AGAGE).[22]
89 GSN station, located on Jeju Island south of the Korean peninsula, is operated by the
90 Kyungpook National University, South Korea. At GSN station, atmospheric mole
91 fractions of HFC-23 are measured every two hours using a cryogenic
92 preconcentration gas chromatograph-mass spectrometry (GC-MS) “Medusa”
93 system.[23, 24] HAT station is situated on a small island at the southwestern edge of the

Japanese archipelago and COI is at the eastern coast of Hokkaido, Japan. Both stations are operated by the National Institute for Environmental Studies, Japan. At HAT and COI stations, atmospheric mole fractions of HFC-23 are measured once per hour using a different technique that also couples a GC-MS system with cryogenic preconcentration.[25, 26] HFC-23 mole fractions at HAT and COI station are originally reported on the NIES-2008 calibration scale; for the sake of consistency, they have been converted to the scale SIO-2005 (the scale of HFC-23 measurements at GSN) by applying a NIES-2008/SIO-2005 conversion ratio of 0.995±0.01.[27] The measured atmospheric HFC-23 mole fraction time series at three stations are shown in Figure 2.

Backward Trajectory Statistics Methods

Modeling Back-trajectories

Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) Model[28] was used in this study to derive 7-day back trajectories from each station. The distances between the HFC-23 sources (almost all in eastern China) and GSN/HAT stations are within a range of 500–1400 kilometers. At these distances, emitted HFC-23 can be transported to the measurement stations within hours to days. Thus, the length of 7-day is sufficient for backward trajectory simulations of HFC-23. The starting locations for the trajectories are at each station with elevations of 20 m, 100 m and 500 m agl, thereby ensuring that the back-trajectory starts in the atmospheric boundary layer[29] (the results of effects of changing release heights on PSCF and CWT

outcomes can be seen below). The HYSPLIT model (http://ready.arl.noaa.gov/HYSPLIT.php), developed by National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory (ARL), is one of the most widely used models for atmospheric trajectory calculations. The meteorological input for the HYSPLIT model running is the Global Data Assimilation System (GDAS) data (ftp://gdas-server.iarc.uaf.edu/gdas1) from the National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) model. The GDAS fields used have a horizontal resolution of 1° and 23 vertical layers. Examples of HYSPLIT back-trajectories are shown in Figure 1b.

PSCF and CWT Analysis

The modeled back-trajectories were used to analyze the potential source regions of HFC-23 in East Asia. For this purpose, the PSCF and CWT methods were employed. The symbols *i* and *j* are the indices of the horizontal grid cell (0.5°×0.5°). The total number of model trajectory endpoints within the grid cell (i, j) is designated as N_{ij}. The number of model trajectory endpoints within the same grid cell that correspond to station HFC-23 mole fractions exceeding a specified threshold (see below the test results of effects of changing threshold criteria on PSCF outcomes) is defined to be M_{ij}. The ratio of M_{ij} to N_{ij}, therefore, represents a unitless fraction of all model trajectory endpoints where the station mole fraction exceeded the specified threshold. The PSCF value for a given grid cell (i, j) is, with the function W described below, defined as

$$PSCF_{ij} = W_{ij} \frac{M_{ij}}{N_{ij}} \quad (1)$$

In the CWT method[6], each grid cell is assigned a weighted concentration, as follows:

$$CWT_{ij} = W_{ij} \frac{1}{\sum_{l=1}^{M} \tau_{ijl}} \sum_{l=1}^{M} C_l \, \tau_{ijl} \quad (2)$$

In Equation 2, CWT_{ij} is CWT value for the grid cell (i, j), l is the index of the trajectory, M is the total number of trajectories, C_l is the enhanced mole fraction of HFC-23 with the corresponding trajectory l, τ_{ijl} is the residence time in the ijth grid cell for the trajectory l.

W_{ij} is an arbitrary weight function to reduce the effect of small values of N_{ij}.[2] W_{ij} was defined as shown in Equation (3), which is consistent with the methods used in previous studies[e.g. 17, 30]. There is an average of 20 endpoints per grid cell for the simulation combining the three release heights (see this calculation in Supplementary Information). Our tests (see Supplementary Information) show that the W_{ij} weight function, even when the threshold of 20 endpoints is doubled in Eq. (3), did not change the PSCF and CWT values at all in the areas where HFC-23 factories are located (because there are significantly more than 20 or 40 endpoints over grid cells in those areas), but affect relatively far away regions, e.g., middle-western China (See Figures S1 and S2).

$$W_{ij} = \begin{cases} 1.00 & N_{ij} > 20 \\ N_{ij}/20 & N_{ij} \leq 20 \end{cases} \quad (3)$$

The higher $PSCF_{ij}$ and C_{ij} values represent in the grid cell (i, j), on average, higher

potential of emission strength. This study used the Geographical Information System (GIS)-based TrajStat software that has been developed by Wang et al.[31] to compute PSCF and CWT analysis.

Inverse Modeling

The inverse modeling method is also used to derive HFC-23 emission source regions in this study. The FLEXPART model was used to calculate the SRR matrix for each station. The backward simulations of FLEXPART were driven by the operational 3-hourly meteorological data at 1°×1° resolution from the European Centre for Medium-Range Weather Forecasts (ECMWF). The inverse modeling method used in this study is the same as described by Fang et al.[27] and Stohl et al.[11]. Briefly, a Bayesian optimization technique is used to estimate emission strength in grid cells (0.5°×0.5°) over the domain influencing the measurement sites. The cost function to be minimized is

$$\boldsymbol{J(x)} = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x_a})^{\mathrm{T}}\mathbf{S}_a^{-1}(\boldsymbol{x} - \boldsymbol{x_a}) + \frac{1}{2}\left(\boldsymbol{y^{obs}} - \mathbf{H}x_a\right)^{\mathrm{T}}\mathbf{S}_{\mathrm{o}}^{-1}\left(\boldsymbol{y^{obs}} - \mathbf{H}x_a\right).$$

We find this minimum by solving $\nabla_x J(\boldsymbol{x}) = 0$, which yields

$$\boldsymbol{x} = \boldsymbol{x_a} + \mathbf{S}_a\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{S}_a\mathbf{H}^{\mathrm{T}} + \mathbf{S}_o)^{-1}(\boldsymbol{y^{obs}} - \mathbf{H}x_a).$$

Here, $\boldsymbol{x}$ is the state vector representing HFC-23 emission strength in each grid cell, $\boldsymbol{x_a}$ is HFC-23 prior emission vector, $\boldsymbol{y^{obs}}$ is HFC-23 measurement vector, $\mathbf{H}$ is the emission sensitivity matrix derived from the FLEXPART backward simulation (see the average emission sensitivities in Figure 3), $\mathbf{S}_a$ is HFC-23 prior emission error

175 covariance matrix, and $\mathbf{S}_o$ is observational error covariance matrix the definition of
176 which is the same as described in detail by Stohl et al..[11] Note that we used flat prior
177 emission of HFC-23 ($\boldsymbol{x}_a$) over the continent (having not included HFC-23 factory
178 location and HFC-23 emission strength; Figure S3), so that the inferred HFC-23
179 emission source map from the inverse modeling method is solely constrained by the
180 observations rather than any information on the location HFC-23 factories.

181 **HFC-23 Factories Information**

182 There are in total twenty-one factories producing HCFC-22 and the byproduct HFC-
183 23 in East Asia in 2010. Eleven HCFC-22 production lines (ten production lines in
184 nine factories in China and one production line in one factory in South Korea) in these
185 twenty-one factories were incinerating HFC-23 in 2010[32] after being required by the
186 program of the United Nations Framework Convention on Climate Change's
187 (UNFCCC) Clean Development Mechanism (CDM) (hollow pentagrams in Figure 4).
188 Negligible amounts of HFC-23 were emitted from these CDM production lines. Three
189 Japanese factories (hollow diamonds in Figure 4) were also incinerating HFC-23,[33]
190 although they did not participate the CDM program in 2010. Factories that have at
191 least one production line without HFC-23 incineration are in China (ten)[21] and North
192 Korea (one)[34], all of which released byproduct HFC-23 directly into the atmosphere in
193 2010 (solid circles in Figure 4). The HFC-23 emissions from North Korea were only
194 ~0.02 Gg/yr in 2010[27, 34], which is much smaller than the emissions of ~6 Gg/yr in
195 China[28]. There were almost no HFC-23 emissions in Mongolia and the Taiwan

196 region.[23] Thus, in the East Asian region, major emissions of HFC-23 were from
197 HCFC-22 factories in China while small amounts of HFC-23 emissions were from
198 North Korea. The detailed information on longitude, latitude, and CDM project
199 enrollment of each factory/production line is provided in Table S1.

200 All back-trajectory statistical and inverse modeling methods in this study are not
201 expected to detect the emission sources of the byproduct HFC-23 factories in western
202 China (there were two HCFC-22 factories with one CMD production line and some
203 non-CDM production lines in 2010), because the factories there (outside of the
204 domain in Figures 4-6) are too far away from the stations for the air mass to carry
205 emission signal to the stations and there were very few air trajectories going from
206 western China to the stations. The emission sensitivities in those grid cells are
207 extremely low ($<10\times10^6$ ppt/(g m^{-2} s^{-1}); "ppt" stands for parts per trillion; Figure 3).
208 Therefore, accurately locating HFC-23 emission sources in western China is outside
209 of the scope of this study.

210 RESULTS AND DISCUSSIONS

211 There have been many studies that have validated trajectory/dispersion models
212 (HYSPLIT, FLEXPART, and so on) including a number that use release of an inert
213 tracer to provide the evaluation data. For example, HYSPLIT, FLEXPART and other
214 models have been evaluated and validated with measurements from the controlled
215 tracer release experiments Cross-Appalachian Tracer Experiment (CAPTEX) and

Across North America Tracer Experiment (ANATEX).[35] However, validation/evaluation of performance of back-trajectory statistical and inverse modeling methods in locating emission source locations is significantly different from validation/evaluation of the trajectory/dispersion models. Some studies[e.g., 15-18] have cross-checked the inferred emission source fields with bottom-up emission information. This study takes advantage of precisely-known locations of HFC-23 sources to evaluate the performance of these three methods in locating the source regions.

Impact of Release Heights

The impact of release heights is one of the most important factors that should be examined before presenting the final results of back-trajectory statistical methods.[17] Figure 4a-c and Figure S4a-c show distributions of CWT and PSCF values, respectively, of HFC-23 emissions sources in East Asia inferred using back-trajectories with release heights of 20 m, 100 m, and 500 m, respectively. Slight differences are observed among the three cases. For example, the 20 m and 100 m cases show higher CWT values along the southern China coast than in the 500 m case. Nevertheless, the CWT value distribution patterns obtained from different release heights are very similar, with high CWT values identified as potential sources of HFC-23 being located in eastern China. Figure 4d shows CWT values calculated using an ensemble of trajectories from these three releases. In Figure 4d, high CWT values are found in eastern China. CWT value distribution pattern in Figure 4d is

consistent with that in Figure 4a-c. Meanwhile, PSCF value distribution patterns obtained from these three release heights are also very similar (Figure S4). Thus, we adopted an ensemble of trajectories from these three releases (reducing the uncertainties associated with the HYSPLIT trajectories from different trajectory release heights) in the PSCF and CWT analysis.

Impact of Threshold Value on PSCF Analysis

A "high mole fraction" threshold value must be chosen for computing the PSCF. Previous studies have used the mean,[e.g., 36] the median[e.g., 2] or the 75th percentile[e.g., 37] of the measurement data as this threshold value. This study tested the values of 30th, 50th, 70th, 90th, 95th and the 98th percentiles of the HFC-23 mole fractions as the threshold values to examine the impact of threshold value on the PSCF results (Figure 5), similar to the comprehensive tests done by Cheng et al.[17]. A wide spread of HFC-23 sources is inferred by the PSCF analysis using the 30th percentile as the threshold value, which does not agree with the actual concentrated HFC-23 sources. As the threshold value increased to 50th and 70th percentiles, high PSCF values became relatively concentrated in eastern China. When the 90th percentile was used, the corresponding PSCF value map shows the potential HFC-23 sources are mainly in eastern China, which is approximately consistent with the actual locations of HFC-23 sources. When 95th and 98th percentiles were used, very few trajectories were identified as emission "signals" and the resulting PSCF values were extremely low. Using the 90th percentile as the threshold value, PSCF analyses produce the best

estimate of HFC-23 source locations, which is consistent with past studies that used the 90th percentile as the threshold value for locating biomass burning sources[17].

PSCF, CWT and Inverse Modeling Results

Figure 6 provides the emission source area information inferred by using the PSCF (with 90th percentile of mole fraction data used as threshold value), CWT and inverse modeling methods. PSCF values (dimensionless) of close to 1.0, representing the potential of high HFC-23 emission, are located in eastern China (Figure 6a), which is in general consistent with the locations of non-CDM byproduct HFC-23 factories. If a lower value is chosen as the threshold value, high PSCF values become spatially diffuse (e.g., in Figure 5a and b), which does not agree at all with HFC-23 sources. The CWT value (in a unit of ppt) distribution map in Figure 6b shows a similar pattern to the PSCF value distribution map in Figure 6a. The PSCF and CWT values in North Korea are small but not negligible, which agrees with the fact that there was a byproduct HFC-23 factory in North Korea and that a small amount of HFC-23 was being emitted from that factory. PSCF and CWT values in Japan are minimal, which is consistent with the fact that almost no HFC-23 was emitted in Japan in 2010.

There are many grid cells with high PSCF and CWT values in northern, middle, and southern China, though non-CDM byproduct HFC-23 factories do not exist in these regions. The misidentification of HFC-23 sources in this region (termed a "ghost source" problem) is likely caused by the climatology, since the air masses passing over ghost source regions had often passed over grid cells with non-CDM

byproduct HFC-23 factories before they reached the measurement stations. The "ghost source" problem originates from the fact that emission sources inferred by the PSCF and CWT methods are combinations of emission sources and preferred transport pathways, and it is hard to distinguish them in the nature of these back-trajectory statistical methods. Maione et al. also identified ghost sources in the wake of real sources in a pseudo-experiment of PSCF method.[4] The "ghost source" problem cannot be fully corrected in the PSCF and CWT methods, though using more stations (ideally placing stations on all sides of all sources, a goal that would be difficult to accomplish in reality) in the PSCF and CWT methods would help reduce the "ghost source" problem.

Figure 6c shows the HFC-23 emission strength (in a unit of $pg/m^2/s$) inferred from the inverse modeling method. Three "hot" areas of HFC-23 emissions were inferred in this inverse modeling case. The locations of three hot areas are quite consistent with the locations of non-CDM byproduct HFC-23 factories. However, only minimal HFC-23 emission strength was inferred for the grid cell of the non-CDM byproduct HFC-23 factory in southern China. This finding is not surprising because a minimal amount of HFC-23 (~0.07 Gg/yr; only ~1% of total China's HFC-23 emissions) was emitted in 2010 in that factory and average emission sensitivity for that grid cell ($\sim 20\times10^6$ ppt/(g m^{-2} s^{-1}); Figure 3d) is too low for the emissions to be well constrained from the three stations. HFC-23 emissions in Japan and South Korea inferred by the inverse modeling method were very small (much smaller than

300 emissions in the three hot areas in China), which is consistent with the reported fact of
301 HFC-23 incineration in these two countries.

302 ## Discussion

303 The analysis above shows that both PSCF and CWT methods did not perform well in
304 accurately locating the emissions source areas. However, PSCF and CWT methods
305 have been widely applied in numerous previous studies. Among the three methods,
306 the inverse modeling method identified most closely the major source areas of HFC-
307 23 emissions. Thus, we suggest that the inverse modeling method should be used
308 more in future studies with the goal of locating emission source regions, and that
309 PSCF and CWT methods should be applied with caution in this kind of study. Note
310 that if the purpose of a study is to identify the mixture of potential sources and
311 transport pathways, beyond the topic of this study, PSCF and CWT methods may be
312 used.

313 Several causes contribute to the poor performance of PSCF and CWT methods in
314 this study. First, contributions of unit emission from each grid cell along a trajectory
315 are equally weighted in calculating the mole fraction at the measurement station,
316 while this is not true in reality. The SRR matrix used in inverse modeling method
317 explicitly shows the respective contribution of unit emission from each grid cell to
318 mole fraction at the receptor. Second, PSCF and CWT values could be biased towards
319 high values for regions which are frequently crossed by trajectories prior to and after
320 the regions with high HFC-23 emissions, and true source areas of high emissions tend

to become spatially diffuse. Third, due to the nature of the PSCF method algorithm, PSCF values may be the same no matter whether the pollutant concentrations at the measurement site are slightly or significantly enhanced, although the CWT method overcomes this shortcoming of the PSCF method. The inverse modeling method overcomes all of these methodological shortcomings very well because it fully accounts for the magnitude of HFC-23 concentrations at the receptor stations.

Another advantage of the inverse modeling method is its ability to determine absolute emission strength (e.g., in a unit of $g/m^2/s$ for the HFC-23 emissions in this study), while the PSCF and CWT methods can only provide relative emission strength potential. The emission field inferred from the inverse modeling method, as opposed to those from the PSCF and CWT methods, could be used to quantitatively cross-check with the bottom-up emission inventory. Thus, the inverse modeling method is well suited for expanded use in future studies.

One limitation of inverse modeling method is that it is computationally expensive to produce the source-receptor-relation matrix as the input of the inverse modeling method. However, back-trajectory statistical methods are computationally fast and they are able to deliver first hints in locating potential emission source areas (note that they are mixed with the preferred transport pathways), despite the limitations of back-trajectory statistical methods. As computational capacity continues to increase quickly, the shortcomings of computationally expensive inverse modeling method becomes much less significant. We suggest that future studies complement with the

342 advantages among the three methods to better understand the emission sources.

ASSOCIATED CONTENT

Supporting Information

Table S1 provides the detailed information on longitude, latitude, and CDM project enrollment of each factory/production line. Figures S1 and S2 show effect of changing the weight function W_{ij} on the resulting PSCF and CWT maps, respectively. Figure S3 shows flat prior emission of HFC-23 in the inverse modeling. Figure S4 shows PSCF inferred HFC-23 emissions sources from PSCF analysis using trajectories from three different release heights. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*e-mail: fangxk@mit.edu

Notes

The authors declare no competing financial interest.

ACKNOLEWGEMENTS

X. Fang and R. G. Prinn are supported by the National Aeronautics and Space Administration (NASA, USA) grants NAG5-12669, NNX07AE89G, NNX11AF17G and NNX16AC98G to MIT. S. Park, S. Li, and the Gosan AGAGE station are

supported by the Korea Meteorological Administration Research and Development Program (KMIPA 2016-6050). Observations at Hateruma Island and Ochiishi were supported by the Global Environment Fund (Ministry of the Environment of Japan). We thank Michael McClellan and Martin Wolf at the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology for language and style edits to this paper.

REFERENCES

(1) Fleming, Z. L.; Monks, P. S.; Manning, A. J. Review: Untangling the influence of air-mass history in interpreting observed atmospheric composition. **2012**, *104–105* (0), 1-39.

(2) Polissar, A. V.; Hopke, P. K.; Paatero, P.; Kaufmann, Y. J.; Hall, D. K.; Bodhaine, B. A.; Dutton, E. G.; Harris, J. M. The aerosol at Barrow, Alaska: long-term trends and source locations. *Atmos. Environ.* **1999**, *33* (16), 2441-2458.

(3) Ashbaugh, L. L. A Statistical Trajectory Technique for Determining Air Pollution Source Regions. *J. Air Waste Manag. Assoc.* **1983**, *33* (11), 1096-1098.

(4) Maione, M.; Giostra, U.; Arduini, J.; Belfiore, L.; Furlani, F.; Geniali, A.; Mangani, G.; Vollmer, M. K.; Reimann, S. Localization of source regions of selected hydrofluorocarbons combining data collected at two European mountain stations. *Sci. Total Environ.* **2008**, *391* (2-3), 232-240.

(5) Cheng, Y.; Lee, S. C.; Ho, K. F.; Wang, Y. Q.; Cao, J. J.; Chow, J. C.; Watson, J. G. Black carbon measurement in a coastal area of south China. *J. Geophys. Res. Atmos.* **2006**, *111* (D12), 12300-12310.

(6) Hsu, Y.-K.; Holsen, T. M.; Hopke, P. K. Comparison of hybrid receptor models to locate PCB sources in Chicago. *Atmos. Environ.* **2003**, *37* (4), 545-562.

(7) Reimann, S.; Schaub, D.; Stemmler, K.; Folini, D.; Hill, M.; Hofer, P.; Buchmann, B.; Simmonds, P. G.; Greally, B. R.; O'Doherty, S. Halogenated greenhouse gases at the Swiss High Alpine Site of Jungfraujoch (3580 m asl): Continuous measurements and their use for regional European source allocation. *J. Geophys. Res. Atmos.* **2004**, *109* (D5), 5307-5318.

(8) Liu, N.; Yu, Y.; He, J.; Zhao, S. Integrated modeling of urban-scale pollutant transport: application in a semi-arid urban valley, northwestern China. *Atmos. Pollut. Res.* **2013**, *4* (3), 306-314.

(9) Stohl, A.; Forster, C.; Frank, A.; Seibert, P.; Wotawa, G. Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. *Atmos. Chem. Phys.* **2005**, *5* (9), 2461-2474.

(10) Jones, A.; Thomson, D.; Hort, M.; Devenish, B., *The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III.* In *Air Pollution Modeling and Its Application XVII*; Borrego, C.; Norman, A.-L., Eds.; Springer US: 2007; pp 580-589.

(11) Stohl, A.; Seibert, P.; Arduini, J.; Eckhardt, S.; Fraser, P.; Greally, B. R.; Lunder, C.; Maione, M.; Mühle, J.; O'Doherty, S.; Prinn, R. G.; Reimann, S.; Saito, T.; Schmidbauer, N.; Simmonds, P. G.; Vollmer, M. K.; Weiss, R. F.; Yokouchi, Y. An analytical inversion method for determining regional and global emissions of greenhouse gases: Sensitivity studies and application to halocarbons. *Atmos. Chem. Phys.* **2009**, *9* (5), 1597-1620.

(12) Cui, Y. Y.; Brioude, J.; McKeen, S. A.; Angevine, W. M.; Kim, S.-W.; Frost, G. J.; Ahmadov, R.; Peischl, J.; Bousserez, N.; Liu, Z.; Ryerson, T. B.; Wofsy, S. C.; Santoni, G. W.; Kort, E. A.; Fischer, M. L.; Trainer, M. Top-down estimate of methane emissions in California using a mesoscale inverse modeling technique: The South Coast Air Basin. *J. Geophys. Res. Atmos.* **2015**, *120* (13), 6698-6711.

(13) Dimitriou, K.; Remoundaki, E.; Mantas, E.; Kassomenos, P. Spatial distribution of source areas of PM2.5 by Concentration Weighted Trajectory (CWT) model applied in PM2.5 concentration and composition data. *Atmos. Environ.* **2015**, *116* (Supplement C), 138-145.

(14) Thompson, R. L.; Patra, P. K.; Chevallier, F.; Maksyutov, S.; Law, R. M.; Ziehn, T.; van der Laan-Luijkx, I. T.; Peters, W.; Ganshin, A.; Zhuravlev, R.; Maki, T.; Nakamura, T.; Shirai, T.; Ishizawa, M.; Saeki, T.; Machida, T.; Poulter, B.; Canadell, J. G.; Ciais, P. Top–down assessment of the Asian carbon budget since the mid 1990s. *Nat. Commun.* **2016**, *7*, 10724.

(15) Scheifinger, H.; Kaiser, A. Validation of trajectory statistical methods. *Atmos. Environ.* **2007**, *41* (39), 8846-8856.

(16) Han, Y. J.; Holsen, T. A.; Hopke, P. K.; Yi, S. M. Comparison between back-trajectory based modeling and Lagrangian backward dispersion modeling for locating sources of reactive gaseous mercury. (vol 39, pg 1715, 2005). *Environ Sci Technol* **2005**, *39* (10), 3887-3887.

(17) Cheng, M. D.; Lin, C. J. Receptor modeling for smoke of 1998 biomass burning in Central America. *J Geophys Res-Atmos* **2001**, *106* (D19), 22871-22886.

(18) Ara Begum, B.; Kim, E.; Jeong, C.-H.; Lee, D.-W.; Hopke, P. K. Evaluation of the potential source contribution function using the 2002 Quebec forest fire episode. **2005**, *39* (20), 3719-3724.

(19) Kabashnikov, V. P.; Chaikovsky, A. P.; Kucsera, T. L.; Metelskaya, N. S. Estimated accuracy of three common trajectory statistical methods. *Atmos. Environ.* **2011**, *45* (31), 5425-5430.

(20) Forster, P.; Ramaswamy, V.; Artaxo, P.; Berntsen, T.; Betts, R.; Fahey, D. W.; Haywood, J.; Lean, J.; Lowe, D. C.; Myhre, G.; Nganga, J.; Prinn, R.; Raga, G.; Schulz, M.; Dorland, R. V., *Changes in Atmospheric Constituents and in Radiative Forcing.* In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, Z.; Avery, K. B.; Tignor, M.; Miller, H. L., Eds.; Cambridge University Press: Cambridge, United Kingdom, 2007; pp 129–234.

(21) Fang, X.; Miller, B. R.; Su, S.; Wu, J.; Zhang, J.; Hu, J. Historical emissions of HFC-23 (CHF3) in China and projections upon policy options by 2050. *Environ. Sci. Technol.* **2014**, *48* (7), 4056-4062.

(22) Prinn, R. G.; Weiss, R. F.; Arduini, J.; Arnold, T.; DeWitt, H. L.; Fraser, P. J.; Ganesan, A. L.; Gasore, J.; Harth, C. M.; Hermansen, O.; Kim, J.; Krummel, P. B.; Li, S.; Loh, Z. M.; Lunder, C. R.; Maione, M.; Manning, A. J.; Miller, B. R.; Mitrevski, B.; Mühle, J.; O'Doherty, S.; Park, S.; Reimann, S.; Rigby, M.; Saito, T.; Salameh, P. K.; Schmidt, R.; Simmonds, P. G.;

Steele, L. P.; Vollmer, M. K.; Wang, R. H.; Yao, B.; Yokouchi, Y.; Young, D.; Zhou, L. History of Chemically and Radiatively Important Atmospheric Gases from the Advanced Global Atmospheric Gases Experiment (AGAGE). *Earth Syst. Sci. Data Discuss.* **2018**, *2018*, 1-39.

(23) Kim, J.; Li, S.; Kim, K. R.; Stohl, A.; Mühle, J.; Kim, S. K.; Park, M. K.; Kang, D. J.; Lee, G.; Harth, C. M.; Salameh, P. K.; Weiss, R. F. Regional atmospheric emissions determined from measurements at Jeju Island, Korea: Halogenated compounds from China. *Geophys. Res. Lett.* **2010**, *37* (12), L12801, 10.1029/2010GL043263.

(24) Miller, B. R.; Weiss, R. F.; Salameh, P. K.; Tanhua, T.; Greally, B. R.; Mühle, J.; Simmonds, P. G. Medusa: A sample preconcentration and GC/MS detector system for in situ measurements of atmospheric trace halocarbons, hydrocarbons, and sulfur compounds. *Anal. Chem.* **2008**, *80* (5), 1536-1545.

(25) Enomoto, T.; Yokouchi, Y.; Izumi, K.; Inagaki, T. Development of an analytical method for atmospheric halocarbons and its application to airborne observation (in Japanese). *J. Jpn. Soc. Atmos. Environ.* **2005**, *40* (1), 1-8.

(26) Yokouchi, Y.; Taguchi, S.; Saito, T.; Tohjima, Y.; Tanimoto, H.; Mukai, H. High frequency measurements of HFCs at a remote site in east Asia and their implications for Chinese emissions. *Geophys. Res. Lett.* **2006**, *33* (21), 814-817.

(27) Fang, X.; Stohl, A.; Yokouchi, Y.; Kim, J.; Li, S.; Saito, T.; Park, S.; Hu, J. Multiannual Top-Down Estimate of HFC-23 Emissions in East Asia. *Environ. Sci. Technol.* **2015**, *49* (7), 4345-4353.

(28) Stein, A. F.; Draxler, R. R.; Rolph, G. D.; Stunder, B. J. B.; Cohen, M. D.; Ngan, F. Noaa's Hysplit Atmospheric Transport and Dispersion Modeling System. *B Am Meteorol Soc* **2015**, *96* (12), 2059-2077.

(29) Dvorska, A.; Lammel, G.; Holoubek, I. Recent trends of persistent organic pollutants in air in central Europe - Air monitoring in combination with air mass trajectory statistics as a tool to study the effectivity of regional chemical policy. *Atmos Environ* **2009**, *43* (6), 1280-1287.

(30) Nicolas, J.; Chiari, M.; Crespo, J.; Galindo, N.; Lucarelli, F.; Nava, S.; Yubero, E. Assessment of potential source regions of PM2.5 components at a southwestern Mediterranean site. *Tellus B* **2011**, *63* (1), 96-106.

(31) Wang, Y. Q.; Zhang, X. Y.; Draxler, R. R. TrajStat: GIS-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data. *Environ. Modell Softw.* **2009**, *24* (8), 938-939.

(32) *Clean Development Mechanism (CDM)*. United Nations Framework Convention on Climate Change: 2014. http://cdm.unfccc.int/Projects/projsearch.html (accessed December 20, 2014).

(33) Greenhouse Gas Inventory Office of Japan (GIO). *National Greenhouse Gas Inventory Report of Japan 2013*, Center for Global Environmental Research (CGER), National Institute for Environmental Studies (NIES), Ministry of the Environment, Japan 2013. http://unfccc.int/files/national_reports/annex_i_ghg_inventories/national_inventories_submissions/application/zip/jpn-2011-nir-26apr.zip (accessed July 24, 2013).

(34) EXECUTIVE COMMITTEE OF THE MULTILATERAL FUND FOR THE IMPLEMENTATION OF THE MONTREAL PROTOCOL. *PROJECT PROPOSAL: DEMOCRATIC PEOPLE'S REPUBLIC OF KOREA*, Montreal, 2012. http://www.multilateralfund.org/68/English/1/6826.pdf (accessed Auguest 1, 2013).

(35) Hegarty, J.; Draxler, R. R.; Stein, A. F.; Brioude, J.; Mountain, M.; Eluszkiewicz, J.; Nehrkorn, T.; Ngan, F.; Andrews, A. Evaluation of Lagrangian Particle Dispersion Models with Measurements from Controlled Tracer Releases. *J Appl Meteorol Clim* **2013**, *52* (12), 2623-2637.

(36) Cheng, M. D.; Hopke, P. K.; Barrie, L.; Rippe, A.; Olson, M.; Landsberger, S. Qualitative Determination of Source Regions of Aerosol in Canadian High Arctic. *Environ Sci Technol* **1993**, *27* (10), 2063-2071.

(37) Gao, N.; Hopke, P. K.; Reid, N. W. Possible sources for some trace elements found in airborne particles and precipitation in Dorset, Ontario. *J Air Waste Manage* **1996**, *46* (11), 1035-1047.

(38) Fang, X.; Shao, M.; Stohl, A.; Zhang, Q.; Zheng, J.; Guo, H.; Wang, C.; Wang, M.; Ou, J.; Thompson, R. L.; Prinn, R. G. Top-down estimates of benzene and toluene emissions in the Pearl River Delta and Hong Kong, China. *Atmos. Chem. Phys.* **2016**, *16* (5), 3369-3382.

Figures

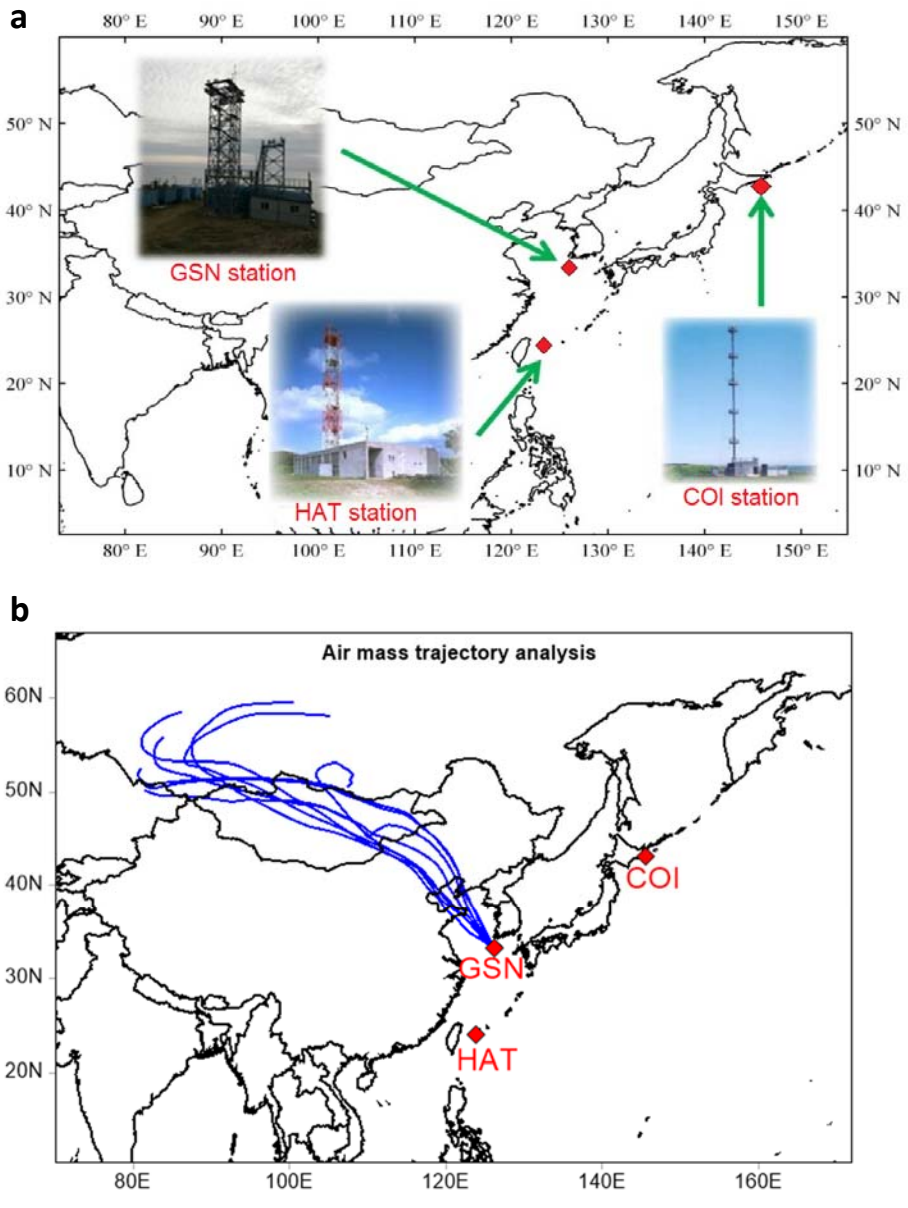
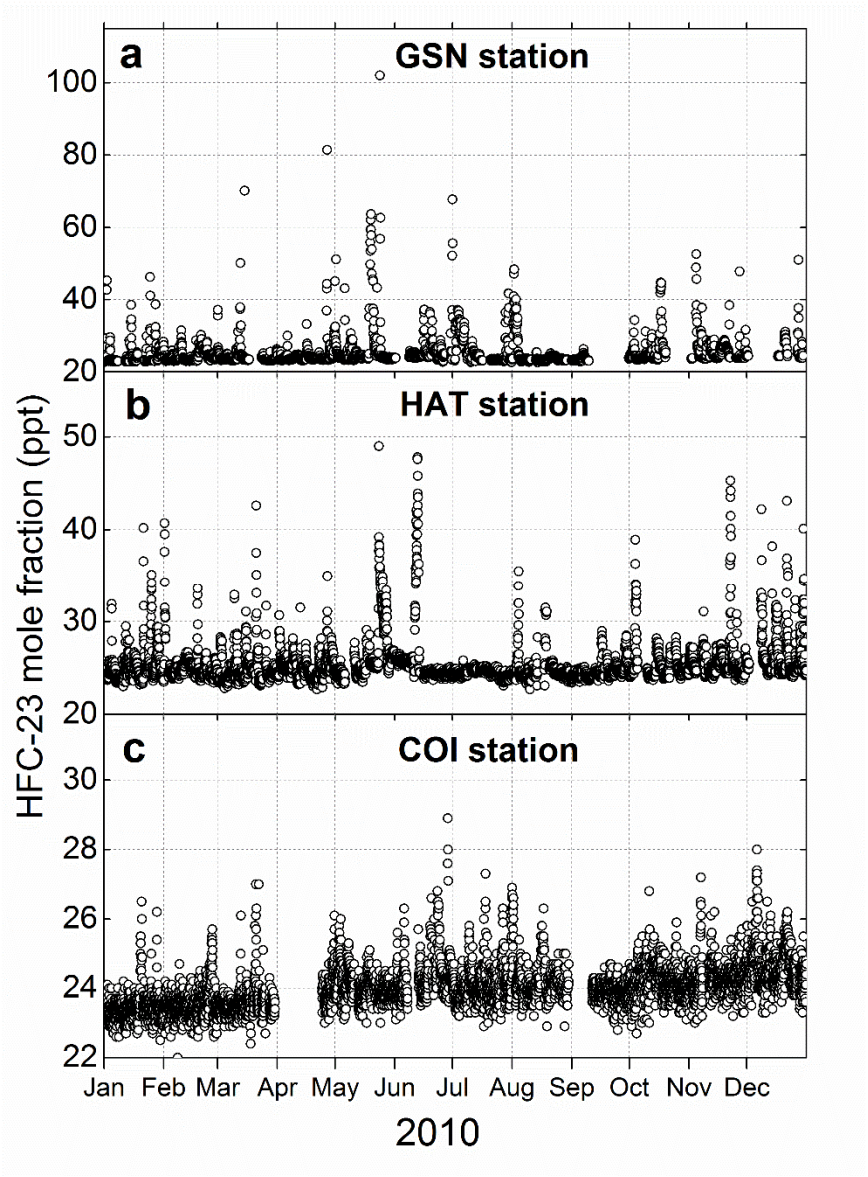


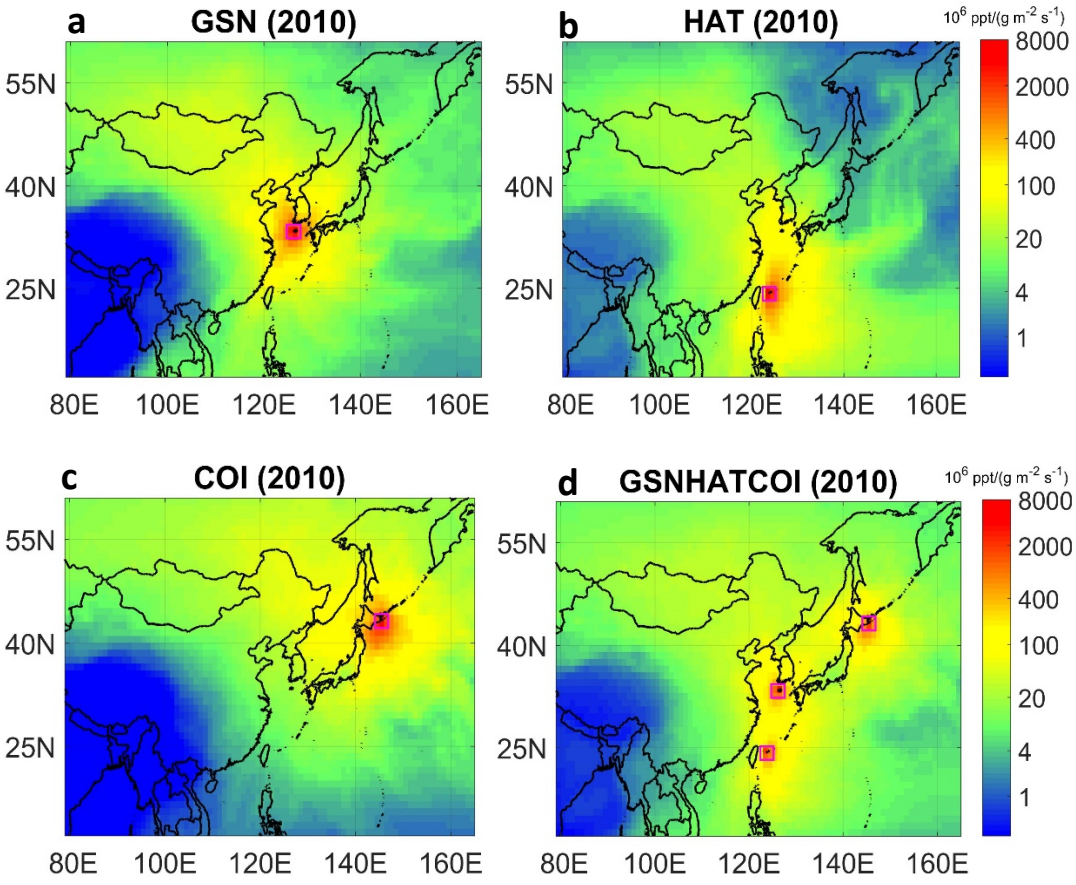
Figure 1. Locations of three HFC-23 measurement stations in East Asia (a) and examples of air mass back-trajectory for the GSN station (b). The red diamonds represent three HFC-23 measurement stations.



501

502 Figure 2. Measured HFC-23 mole fractions at (a) GSN, (b) HAT, and (c) COI stations in
503 East Asia. Note that the y scales are different in three subplots to illustrate the
504 magnitude of HFC-23 mole fractions at each station.

505



506

507 Figure 3. Average emission sensitivity from FLEXPART simulations for (a) GSN, (b) HAT,

508 (c) COI and (d) GSNHATCOI stations for the year 2010. The purple squares represent the

509 HFC-23 measurement stations.

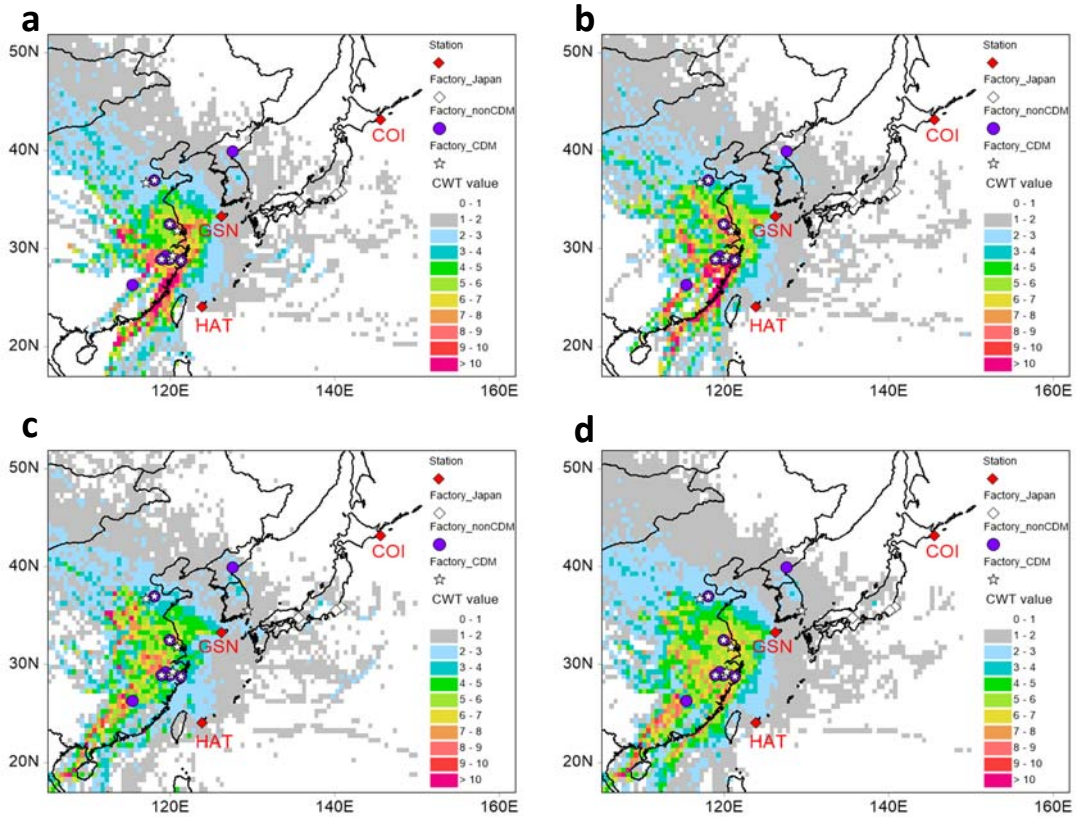


Figure 4. CWT distribution of HFC-23 emissions sources in East Asia determined using release heights of 20 m agl (a), 100 m agl (b) and 500 m agl (c). d are results from CWT analysis using all back-trajectories from the three release heights. The red diamonds represent three HFC-23 measurement stations.

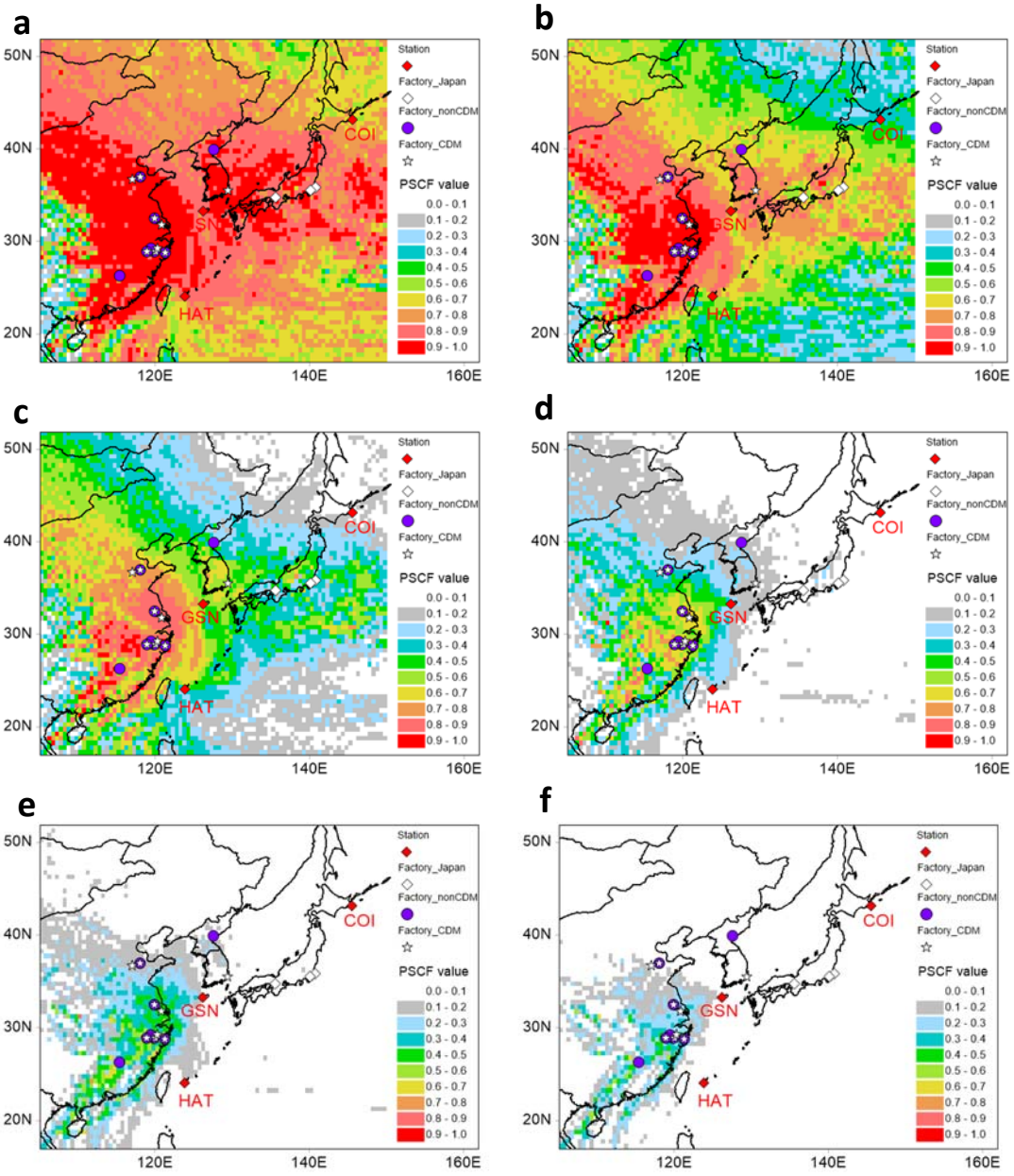


Figure 5. Effect of changing the mole fraction threshold value on the resulting PSCF maps using 30th (a), 50th (b), 70th (c), 90th (d), 95th (e) and 98th (f) percentiles of the HFC-23 mole fractions as the threshold value in PSCF calculation. Back-trajectories from all three release heights (20 m, 50 m and 500 m agl) were used in PSCF calculations.

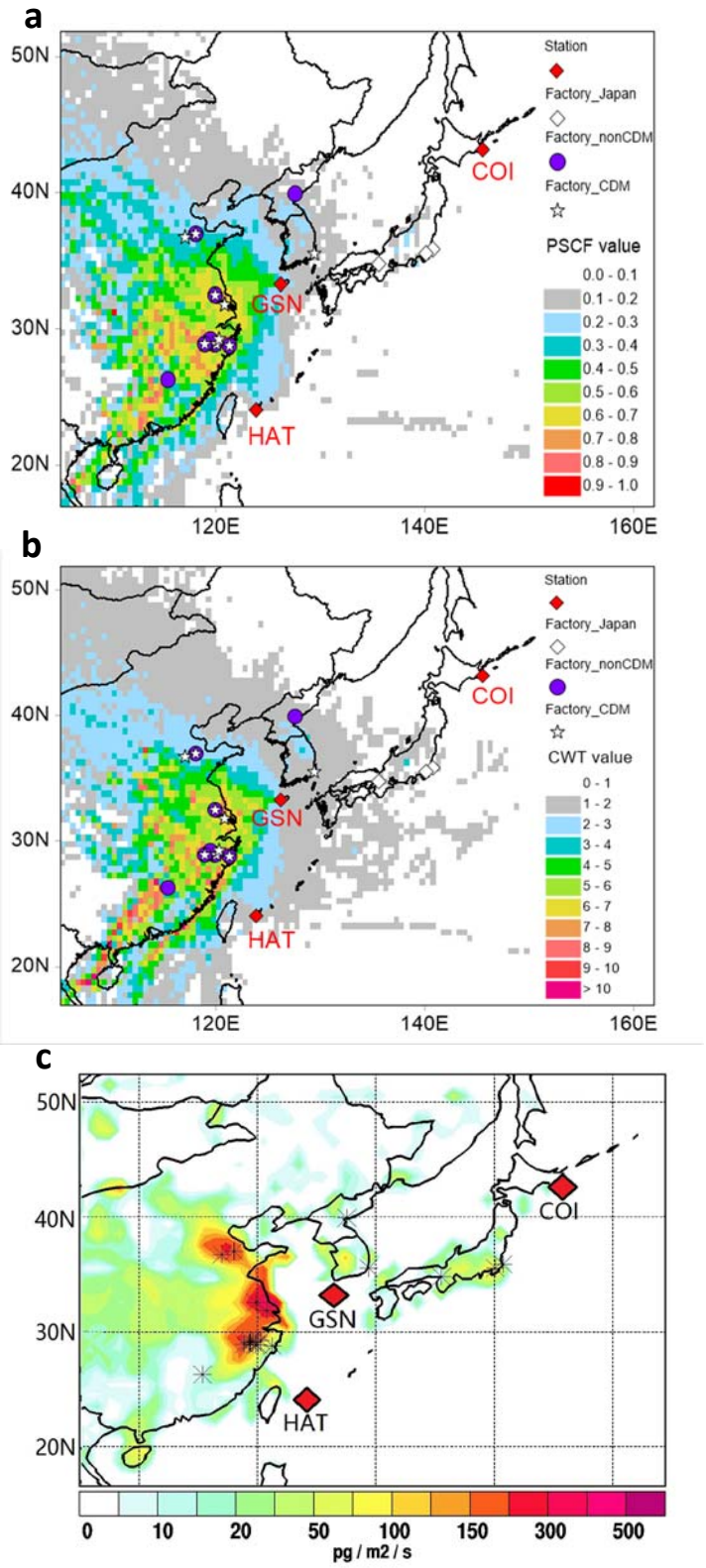


Figure 6. The PSCF (a), CWT (b) and inverse modeling (c) results for HFC-23 emissions in East Asia for the year 2010. The PSCF values were calculated with 90th percentile of mole fraction data used as threshold value. The red diamonds represent three HFC-23 measurement stations. In (c), the asterisks represent byproduct HFC-23 factories including CDM and non-CDM.

Performance of back-trajectory statistical methods and inverse modeling method in locating emission sources

Xuekun Fang[*,†,‡], Takuya Saito[§], Sunyoung Park[∇], Shanlan Li[∇], Yoko Yokouchi[§], Ronald G. Prinn[†].

[†]Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[‡]College of Environmental Sciences and Engineering, Peking University, Beijing, China

[§]National Institute for Environmental Studies, Tsukuba, Japan

[∇]Department of Oceanography, Kyungpook National University, Daegu, South Korea

Tests of Weight Function

The total number of individual observations in 2010 at all three stations was 6993. The back-trajectories were initialized at 3 release heights. The number of endpoints for each trajectory is 56 (7days*8/day). Thus, the total number of endpoints for the whole-year period in this study is 6993*3*56=1174824. The trajectories approximately cover a domain of 0°–180° (longitude) and 0°–80° (latitude). Thus, the total number of grid cells is 57600 (180*80*4) in the geophysical regions covered by the trajectories. Thus, there is an expected average of 20 endpoints per grid cell.

$$W_{ij} = \begin{cases} 1.00 & N_{ij} > 20 \\ N_{ij}/20 & N_{ij} \leq 20 \end{cases}$$

We examined the effect of changing weight function on the resulting PSCF (Figure S1) and CWT (Figure S2) maps using no weight function, 20 as the break point and 40 as the break point in the weight functions. See the equation below for the weight function using 40 as the break point. 90th percentile of the HFC-23 measurement data is used as the threshold value. Results show that the W_{ij} weight function does not change the PSCF values in the areas where HFC-23 factories are located in (because there are significantly more than 20 endpoints over grid cells in the areas), but affect relatively far away regions, e.g., middle-western China. The same W_{ij} weight function was applied in the CWT calculation, and we arrived at the same conclusion of the impact of changing W_{ij} weight function on the CWT map.

29 $$W_{ij} = \begin{cases} 1.00 & N_{ij} > 40 \\ N_{ij}/40 & N_{ij} \leq 40 \end{cases}$$

30 If only one release height is considered, the average number of endpoints per grid
31 cell is about 7. Thus we used 7 as the break point in the weight function in the PSCF
32 and CWT calculations if trajectories from only one release height were used.

33 $$W_{ij} = \begin{cases} 1.00 & N_{ij} > 7 \\ N_{ij}/7 & N_{ij} \leq 7 \end{cases}$$

Tables

Table S1. Information on longitude, latitude, and CDM project enrollment for each byproduct HFC-23 factory (HCFC-22 production factory). Note that "CDM" and "non-CDM" together in a row represents that there are more than one production lines in the corresponding HCFC-22 factory and that some of production lines participated the CDM incineration program. "CDM" for a factory means insignificant emissions of HFC-23; "non-CDM" for a factory means substantial emissions of HFC-23.

Country	Factory ID	Longitude	Latitude	Information on enrollment of CDM	
China	1	119.9	32.5	CDM	non-CDM
	2	120.8	31.8	CDM	
	3	120.8	31.8	CDM	
	4	117.0	36.7	CDM	
	5	118.0	37.0	CDM	non-CDM
	6	119.6	29.1		non-CDM
	7	119.8	28.9		non-CDM
	8	120.0	28.9	CDM	
	9	119.4	29.1		non-CDM
	10	119.4	29.2		non-CDM
	11	118.9	28.9	CDM	non-CDM
	12	121.3	28.8	CDM	
	13	120.3	29.3	CDM	
	14	104.9	29.2	CDM	non-CDM
	15	104.8	29.4		non-CDM
	16	115.4	26.3		non-CDM
South Korea	17	129.4	35.5	CDM	
North Korea	18	127.5	39.9		non-CDM
Japan	20	140.7	35.8		non-CDM but incinerating
	21	135.6	34.8		non-CDM but incinerating
	22	140.1	35.5		non-CDM but incinerating

Figures

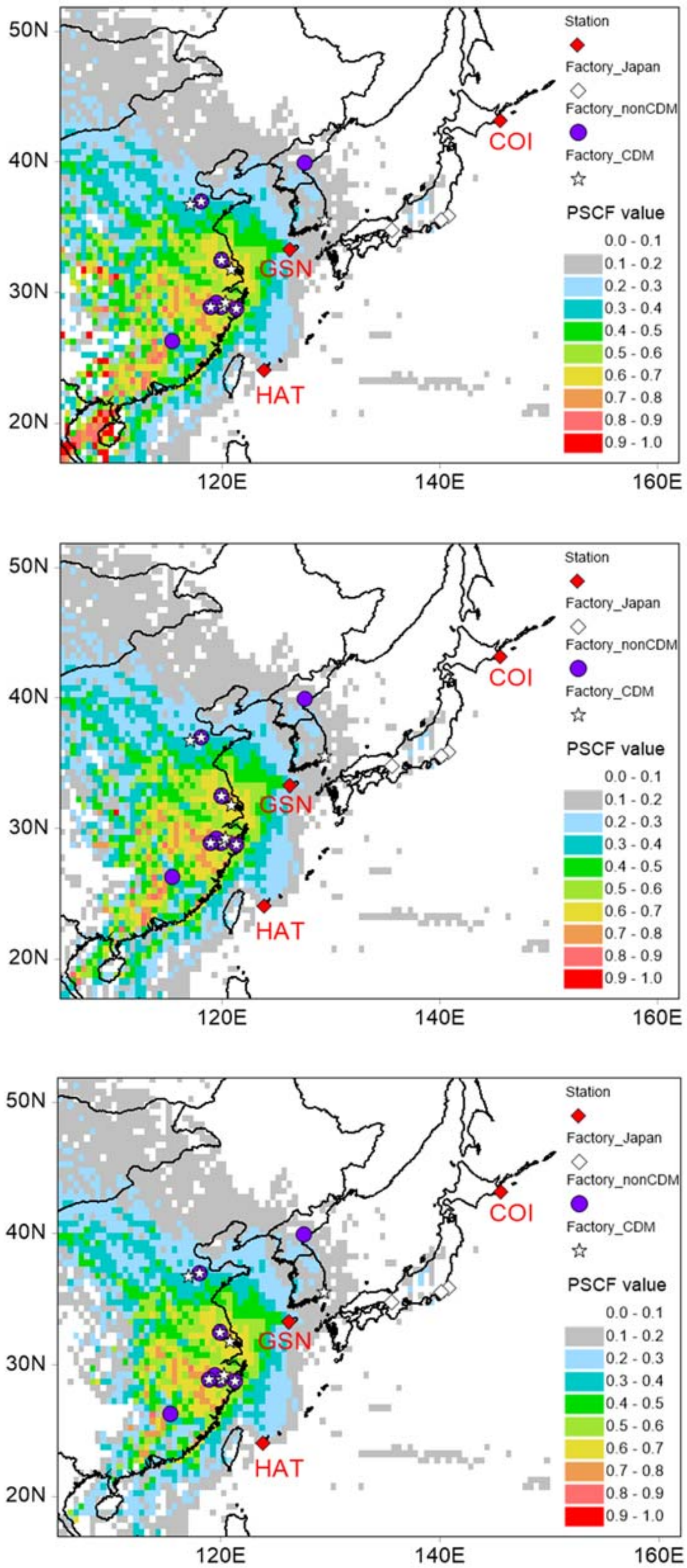


Figure S1. Effect of changing the weight function on the resulting PSCF maps using (a) no weight function, (b) 20 as the break point and (c) 40 as the break point in the weight function. 90^{th} percentile is used as the mole fraction threshold value.

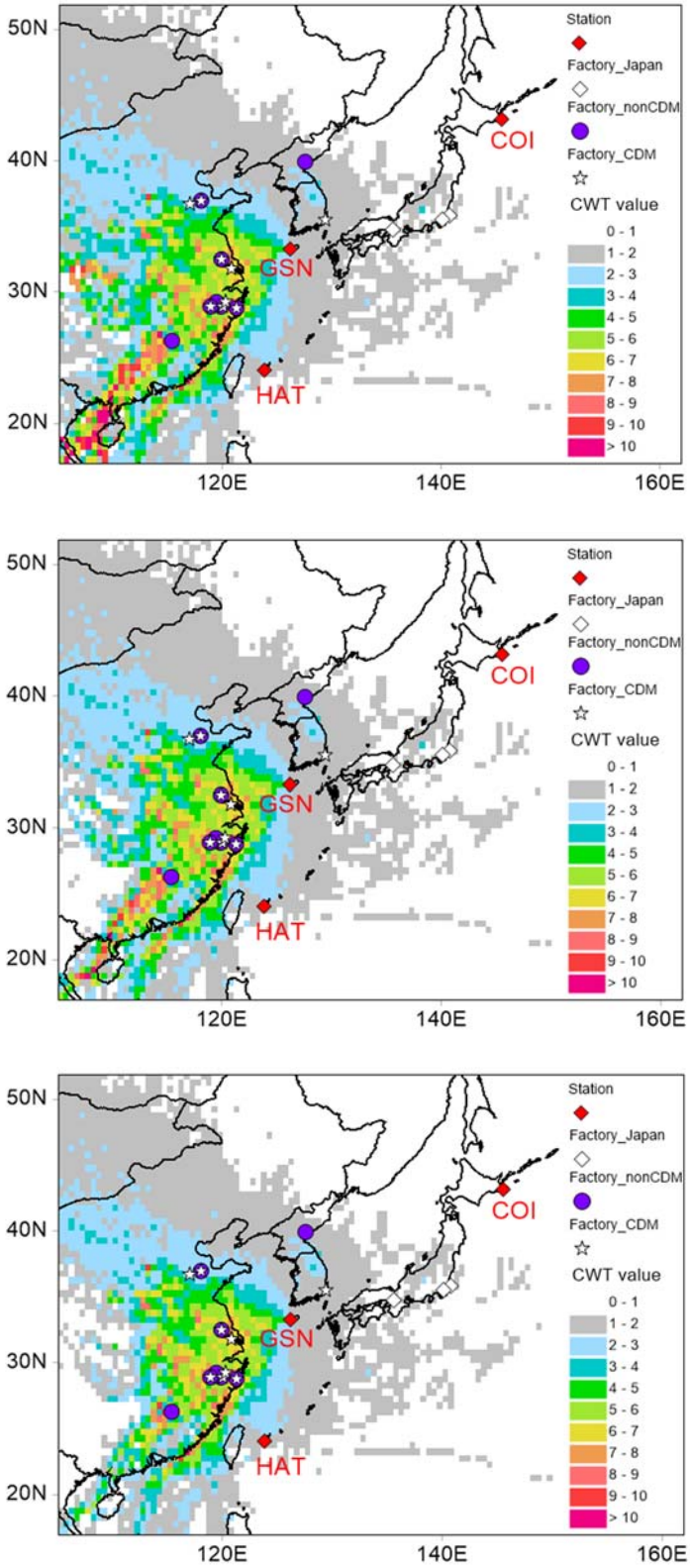


Figure S2. Effect of changing the weight function on the resulting CWT maps using (a) no weight function, (b) 20 as the break point and (c) 40 as the break point in the weight function. 90th percentile is used as the mole fraction threshold value.

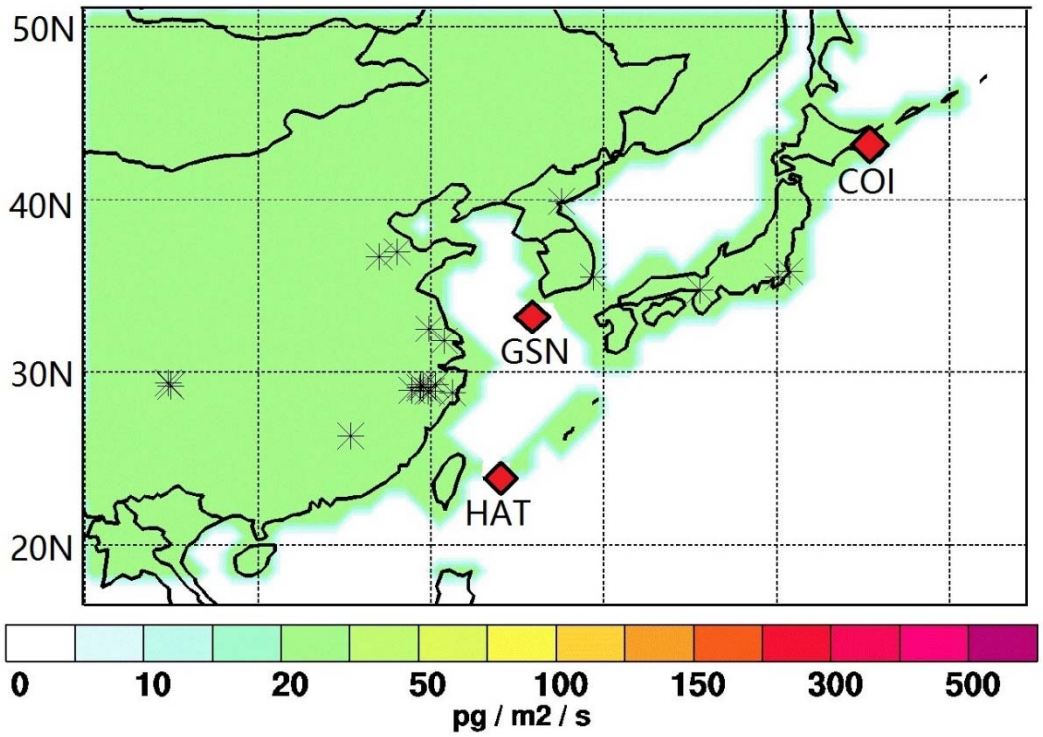


Figure S3. Fat prior emission of HFC-23 in the inverse modeling. The red diamonds represent three HFC-23 measurement stations. The asterisks represent byproduct HFC-23 factories including CDM and non-CDM.

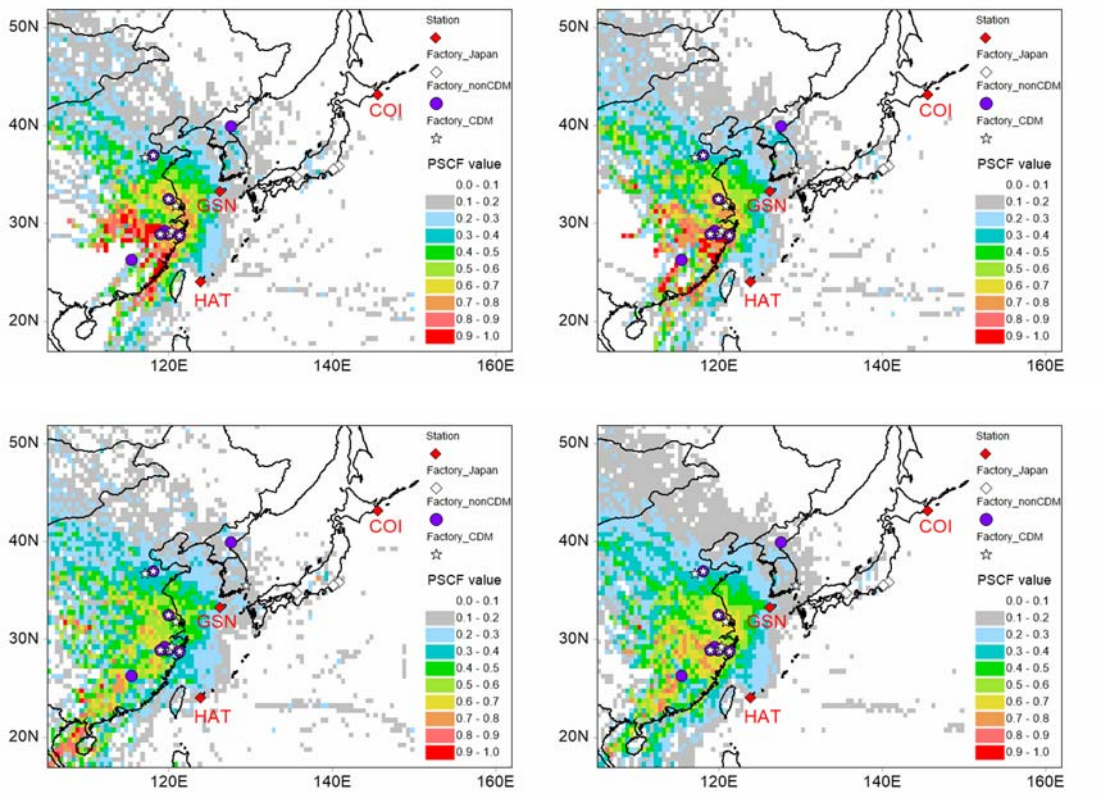


Figure S4. PSCF distribution of HFC-23 emissions sources in East Asia determined using release heights of 20 m agl (a), 100 m agl (b) and 500 m agl (c). d are results from CWT analysis using all back-trajectories from the three release heights. The red diamonds represent three HFC-23 measurement stations.