High-resolution (0.05°×0.05°) NOₓ emissions in the Yangtze River Delta inferred from OMI

Hao Kong[1], Jintai Lin[1], Ruixiong Zhang[1*], Mengyao Liu[1], Hongjian Weng[1], Ruijing Ni[1], Lulu Chen[1], Jingxu Wang[1], Qiang Zhang[2]

5 [1]Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China

[2]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China

Correspondence to: Jintai Lin (linjt@pku.edu.cn)

10 [*] Now at: School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Emission datasets of nitrogen oxides (NOₓ) at high horizontal resolutions (e.g., $0.05° \times 0.05°$) are crucial for understanding human influences at fine scales, air quality studies, and pollution control. Yet high-resolution emission data are often lacking or contain large uncertainties especially for the developing

15 regions. Taking advantage of long-term satellite measurements of nitrogen dioxide (NO₂), here we develop a computationally efficient method to inverting NOₓ emissions in major urban areas at the $0.05° \times 0.05°$ resolution. The inversion accounts for the nonlinear effects of horizontal transport, chemical loss, and deposition. We construct a 2-dimensional Peking University High-resolution Lifetime-Emission-Transport (PHLET) model, its adjoint model (PHLET-A), and a Satellite Conversion Metrix

20 approach to relate emissions, simulated NO₂, and satellite NO₂ data. The inversion method is applied to summer months of 2012–2016 in the Yangtze River Delta area (YRD, 118°E-123°E, 29°N-34°N), a major

polluted region of China, using the POMINO NO_2 vertical column density product retrieved from the Ozone Monitoring Instrument. A systematic analysis of inversion errors is performed, including using an Observing System Simulation Experiment-like test. Across the YRD area, the inverted summer average emission ranges from 0 to 12.0 kg km^{-2} h^{-1}, and the lifetime (due to chemical loss and deposition) from

5   1.4 to 3.6 h. Our inverted emission dataset reveals fine-scale spatial information tied to nighttime light, population density, road network, and maritime shipping. Many of the inverted fine-scale emission features are not well represented or not included in the widely used Multi-scale Emissions Inventory of China. Our inversion method can be applied to other regions and other satellite sensors such as the TROPOspheric Monitoring Instrument.

10   **1.   Introduction**

Nitrogen oxides (NO_x = NO + NO_2) are a main precursor of particulate matter, ozone, and other atmospheric pollutants. NO_x strongly influence the atmospheric oxidative capacity, affect the climate, and are toxic to many organisms. NO_x are emitted from natural (soil, biomass burning, and lightning) and anthropogenic (transportation, energy, industry, and residential) sources (Lin, 2012). Over the past decade,

15   China has experienced rapid growth in the Gross Domestic Product (GDP, by 8.3% a^{-1} on average from 2008 to 2017), fossil fuel consumption (by 5.5% a^{-1} from 2007 to 2015), and urbanization (National Bureau of Statistics of China, http://data.stats.gov.cn/). These socioeconomic changes have been accompanied by a rapid change in NO_x emissions in the urban and surrounding areas. With the large and continuously increasing urban population and motor vehicles, NO_x pollution is particularly severe in large

20   cities such as Beijing and Shanghai (Barnes and Rudziński, 2013;Lin et al., 2016). Many coastal cities like Shanghai have also experienced enormous growth in the shipping business and resulting pollution along with the growing volume of global economic trade. Emissions from seaborne transport play an increasingly important role in the global air pollution (Fu et al., 2017). Understanding the urban pollution and its environmental impacts requires accurate quantitative knowledge on NO_x emissions at a very high

25   horizontal resolution, which is still lacking especially for the developing countries.

Gridded bottom-up emission inventories typically use spatial proxies (like population and GDP) to allocate provincial-level emission values to individual locations (Zhao et al., 2011;Janssens-Maenhout et al., 2015), which may lead to large uncertainties at high resolutions (Zhao et al., 2011;Zheng et al., 2017;Geng et al., 2017). Top-down inversion using satellite retrieval products of tropospheric vertical

5   column densities (VCDs) of nitrogen dioxide (NO_2) is a widely used independent estimate of NO_x emissions(Martin et al., 2003;Stavrakou et al., 2008;Lin et al., 2010;Mijling and R., 2012;Gu et al., 2014;Beirle et al., 2015;Miyazaki et al., 2016;Ding et al., 2017a). The traditional top-down methods based on local mass balance (LMB) or its variants assume a weak effect of horizontal transport (Martin et al., 2003;Lamsal et al., 2011;Lin, 2012;Gu et al., 2014;Boersma et al., 2015). These algorithms work

10   relatively well at low resolutions (> 50 km) given the relatively short lifetime of NO_x (hours to 1 day), but may introduce large uncertainties when applied to higher resolutions – for example, emissions in the rural-urban fringe zone cannot be identified accurately. The Adjoint Model and Kaman Filter methods better account for horizontal transport, although their applicability is limited by expensive computational costs (Stavrakou et al., 2008;Mijling and R., 2012;Miyazaki et al., 2016;Qu et al., 2017;Ding et al., 2017a).

15   These more sophisticated methods are typically applied to limited time periods, limited spatial domains, and/or low horizontal resolutions (Gu et al., 2014;Ding et al., 2017b;Miyazaki et al., 2012;Stavrakou et al., 2008;Qu et al., 2017;Huang et al., 2014). The LMB, Adjoint Model and Kaman Filter approaches normally use 3-dimentional chemical transport models (CTMs) to relate emissions to VCDs. With a few exceptions (Lin et al., 2012;Stavrakou et al., 2013), the uncertainties in CTMs are not often assessed

20   comprehensively due to expensive computational costs. A computationally low-cost method for space-based high-resolution NO_x emission inversion will be helpful for understanding the urban pollution and its trends and variability. This is particularly true with the advent of fine spatial resolution satellite sensors such as the TROPOspheric Monitoring Instrument (TROPOMI).

This study presents a computationally low-cost space-based top-down approach to constructing high-

25   resolution NO_x emission inventory for the urban and surrounding areas. The approach is applied to the Yangtze River Delta (YRD) area (118°E-123°E, 29°N-34°N, which includes Shanghai, Nanjing,

Hangzhou and other 13 cities) on a 0.05°×0.05° grid, using the POMINO NO_2 VCD data retrieved from the Ozone Monitoring Instrument (OMI). We derive the average NO_x emissions for the summer months (June, July, and August) of 2012–2016. We construct a 2-dimensional (2-D) Peking University High-resolution Lifetime-Emission-Transport (PHLET) model and its adjoint model (PHLET-A) to facilitate

5   the emission inversion. The concentration-dependent lifetime of NO_x is estimated as well in order to account for the nonlinear NO_x chemistry.

Section 2 presents the data and method for high-resolution NO_x emission inversion. Inversion uncertainties are analyzed explicitly. We also test the inversion method by applying it to the NO_2 VCDs simulated by the GEOS-Chem CTM, in a manner similar to the Observing System Simulation Experiment

10   (OSSE). Section 3 presents the top-down estimate of NO_x emissions, lifetimes, and their uncertainties. Comparison with the Multi-scale Emissions Inventory of China (MEIC) is also provided. Section 4 concludes the study.

2. Data and Method

2.1. A general framework to retrieve NO_x emissions at a high resolution

15   The high-resolution NO_x emission retrieval framework consists of multiple steps. First, the POMINO NO_2 VCD data over summer 2012–2016 are averaged on a 0.05°×0.05° grid, using a special oversampling technique that preserves the finest spatial information possible. Second, the PHLET model is constructed to simulate the local net source (emission – loss) and horizontal transport of NO_2 VCD on the 0.05°×0.05° grid. A Satellite Conversion Matrix (SCM) approach is used for spatial conversion between the PHLET

20   grid space and the satellite pixel space. Third, the PHLET-A adjoint model is constructed to, together with PHLET, invert the local net source from satellite NO_2 VCD data. Finally, the emission and lifetime of NO_x at each grid cell is derived from the inverted local net source term, by assuming a fixed formula within the small study domain for the nonlinear relationship between lifetime and VCD. A rigorous error analysis for the framework and models is also conducted, including using an OSSE-like test.

Our inversion method explicitly accounts for horizontal transport and the nonlinear relationship between NO_x emissions, lifetimes, and NO_2 VCDs. With a few reasonable assumptions, the method is computationally efficient, suitable for speedily conducting high-resolution emission inversion in multiple areas and across a long time period. In the case of YRD here, with one computational core (Intel® Xeon®

5 Gold 6130 CPU @ 2.10GHz), the inversion calculation takes about 36 hours after necessary input data are prepared.

2.2. Tropospheric NO₂ VCDs retrieved from OMI

OMI is a UV/VIS nadir solar backscatter spectrometer on board the Aura satellite. OMI provides daily global coverage. Each complete swath of OMI consists of 60 ground pixels, the sizes of which increase

10 from 13 km × 24 km at nadir to about 40 km × 150 km at the swath edge in accordance to the view zenith angle (VZA) from 0° to 57° (de Graaf et al., 2016).

We use Level-2 tropospheric NO_2 VCD data from the POMINO product (Lin et al., 2014;Lin et al., 2015). As described in detail in Lin et al. (2014; 2015), POMINO is an OMI-based regional NO_2 product that includes a number of important features. Briefly, POMINO adopts the tropospheric slant column density

15 (SCD) data from DONIMO v2 and conducts an improved calculation of tropospheric air mass factors (AMFs) and VCDs (i.e., VCD = SCD / AMF) (Boersma et al., 2011). Key features of the POMINO algorithm include explicit representation of aerosol scattering and absorption (by combining aerosol data from daily nested GEOS-Chem (at 0.3125° long. × 0.25° lat.) simulations and monthly MODIS/Aqua aerosol optical depth (AOD) data), explicit representation of the angular dependence of surface reflection,

20 high-resolution NO_2 profiles from GEOS-Chem (at 0.3125° long. × 0.25° lat.), consistent retrievals of clouds (a prerequisite for the NO_2 retrieval) and NO_2, and use of a parallelized, LIDORT-driven AMFv6 package. POMINO NO_2 VCDs are consistent with ground-based MAX-DOAS data (Liu et al., 2018b).

To better relate NO_x emissions to NO_2 VCDs at the 0.05°×0.05° resolution, we only employ the NO_2 data in summer (June, July, and August), in which season the lifetime of NO_2 is the shortest (a few hours). We

25 combine data over 2012–2016 to increase the sample size. The change in NO_2 VCD from June to August

is relatively small, reducing the effect of intra-seasonal variability when inverting NO_x emissions from summer mean NO_2 VCDs. We screen out the 30 outer pixels with VZA larger than 30° (cross-track width larger than 36 km) that greatly smear the spatial gradient of NO_2, pixels with cloud radiance fraction exceeding 50%, and pixels with AOD larger than 3 (i.e., when the aerosol data used in the NO_2 retrieval

5   are unreliable). We also exclude data with raw anomaly problems (http://projects.knmi.nl/omi/research/product/rowanomaly-background.php). We then convert the pixel-specific Level-2 data to the 0.05°×0.05° grid.

To convert from the satellite pixel space to the 0.05°×0.05° grid space, we use an oversampling method that employ satellite data on multiple days to enhance the horizontal resolution(Zhang et al., 2014). For

10   each 0.05°×0.05° grid cell, we average all pixels covering the grid cell from all valid days, using area-based weighting. The oversampling approach takes advantage of the fact that the exact footprint of the OMI instrument slightly changes from one day to another, so does the exact footprint of a satellite pixel at a given VZA. Thus, sampling from multiple days increases the horizontal resolution of data. Our oversampling approach is different from Fioletov et al., who filled a grid cell with data from pixels within

15   a certain distance (e.g., 30 km), which implied a spatial smoothing (Fioletov et al., 2011;Krotkov et al., 2016;Sun et al., 2018).

For purpose of emission inversion, we assume that the error of VCD at a satellite pixel (σ_p) contains an absolute error of 0.5×10^{15} molecules cm^{-2} and a relative error of 30% (Lin et al., 2010;Boersma et al., 2011;Lin et al., 2015;Beirle et al., 2011). We further add in quadrature an additional error (σ_g) when a

20   satellite pixel is projected to the grid cells at a finer resolution; this error is important in the urban-rural fringe zone. For a given grid cell, σ_g is set to be 15% of the standard deviation of VCDs at its surrounding eight grid cells. Sampling over multiple days reduces the random error by a factor of $s = \left(\sqrt{(1-c)/n + c}\right)$, where c represents the fraction of systematic error (assumed to be 50%) and n the number of days with valid data (Eskes et al., 2003;Miyazaki et al., 2012). Thus, the total error for the

25   temporally averaged VCD at a given grid cell is $\sigma_s = \sqrt{(\sigma_p^2 + \sigma_g^2)} \cdot s$.

2.3. The PHLET model simulation

We construct the PHLET model on the $0.05° \times 0.05°$ grid to interpret the relationship between local net source (emissions – loss), horizontal transport, and VCD of NO_2 in a 2-D gridded space (Eq. (1)) in the sense of long-time average. PHLET simulates the horizontal transport of NO_x through a time averaged

5   advection process and an "effective" diffusion process, which represents the residual from the temporally averaged advection. The model assumes a horizontally homogeneous vertical shape of NO_2 concentrations, and that NO_2 is concentrated near the surface (Beirle et al., 2011). The assumption is implicitly used in many previous studies for polluted areas (Liu et al., 2016;Beirle et al., 2011). The corresponding uncertainty in the modeled NO_2 VCDs is set as 15% (Lin et al., 2014;Boersma et al., 2011).

10 Although lightning emissions, biomass burning emissions, aircraft emissions, and the vertical transport can lead to NO_2 at higher altitudes over the YRD area, the amount of NO_2 aloft is much smaller than NO_2 near the ground due to large anthropogenic pollution emitted from the ground (Lin, 2012). We regard NO_2 aloft as the regional background, and do not include it in Eq. 1 – as such, PHLET accounts for anthropogenic and soil NO_x emitted from the surface but not other types of emission sources. [For NO_x

15 near the ground in the YRD area, the contribution of lightning, biomass burning, and aircraft emissions is negligible compared to the major anthropogenic (energy, industry, transportation, residential) sources.] To ensure the consistency between PHLET and OMI NO_2 data, we assume the background value to be half of the minimum OMI NO_2 VCD among all grid cells (i.e. 0.54×10^{15} molecules cm^{-2}), and then subtract the background value from the gridded OMI data when comparing with PHLET simulations. The

20 corresponding uncertainty in the modeled NO_2 VCDs is set as 5%.

Equation (1) presents the governing equation in PHLET:

$$\frac{\partial C(x,y)}{\partial t} = r \cdot L(x,y) - \boldsymbol{V}(x,y) \cdot \nabla C(x,y) + \nabla \cdot (\boldsymbol{K}(x,y) \cdot \nabla C(x,y)) \qquad (1)$$

$C(x,y)$ represents the tropospheric NO_2 VCD (in molecules cm^{-2}) with the regional background removed. The value of $C(x,y)$ after the model reaches equilibrium ($\frac{\partial C(x,y)}{\partial t} = 0$ for every x and y) gives the

distribution of NO$_2$ VCDs to be compared with the gridded, background value excluded OMI data. As such, we assume steady state of NO$_2$ in the PHLET model, and assign an uncertainty of 15% to simulated $C(x, y)$ to account for the possibility that NO$_2$ may not be fully in steady state. Also, combining data from multiple years to derive an averaged NO$_2$ distribution for simulation (rather than conducting the

5   simulations for individual years and months) leads to an additional uncertainty, which is set to be 10%.

$L(x, y)$ represents the local net source term (in molecules cm^{-2} s^{-1}, equivalent to 2.63×10^{-12} kg km^{-2} h^{-1}) combining the effects of emissions (anthropogenic + soil), deposition, and chemistry of NO$_x$, and r the ratio of NO$_2$ over NO$_x$ concentration. The daytime NO$_x$ chemical system reaches equilibrium rapidly and r varies little (Beirle et al., 2011;Valin et al., 2013). We set r to be 0.76 with an uncertainty of 10%

10   (Seinfeld et al., 2006;Beirle et al., 2011). $L(x, y)$ can be separated into an emission term and a loss term:

$$L(x, y) = E(x, y) - \frac{C(x,y)}{r \cdot \tau(x,y)} \qquad (2)$$

where $E(x, y)$ denotes the gridded emission of NO$_x$, and $\tau(x, y)$ the lifetime of NO$_2$ associated with deposition and chemical loss.

$V(x, y) \cdot \nabla C(x, y)$ represents the time averaged advection term. $V(x, y)$ denotes the mean wind vector

15   (in m s^{-1}) averaged over summer 2012–2016. To solve the advection process, we follow the Walcek scheme (Walcek, 2000); the SHASTA scheme (Schere, 1983) leads to similar results.

$\nabla \cdot (\mathbf{K}(x, y) \cdot \nabla C(x, y))$ represents the diffusion term, where $\mathbf{K}(x, y)$ denotes the "effective" diffusion coefficient tensor (in m^2 s^{-1}). The diffusion term accounts for transport by the residual winds deviating from the temporally averaged wind vector $V(x, y)$. Appendix A shows how to determine the diffusion

20   coefficient and solve the diffusion process.

The discrete form of PHLET is set on the $0.05° \times 0.05°$ grid. To run PHLET, the NO$_2$ VCDs at the domain edges, as the lateral boundary conditions (LBCs), are set as the corresponding OMI NO$_2$ VCDs. For initial conditions, the VCD at each grid cell inside the domain boundaries is set to be zero, and the

local net source term is assigned to be a low value (0.02 kg km^{-2} h^{-1}, or 7.60×10^9 molecules cm^{-2} s^{-1}). The horizontal distribution of modeled NO$_2$ VCDs at equilibrium does not depend on the initial conditions. At equilibrium, the domain average of modeled $L(x,y)$ reaches zero, become there are no horizontal fluxes into or out of the domain boundaries.

5   For horizontal transport, we use 3-hourly wind fields from the European Centre for Medium-range Weather Forecasts (ECMWF) ERA5 dataset (https://confluence.ecmwf.int//display/CKB/ERA5+data+documentation; last access: 2018/7/2). The resolution of raw ERA5 data is 0.28125° on the reduced Gaussian grid, which is regridded to 0.05° × 0.05° by using the online program offered by ECMWF (see Fig. A1). We adopt the mean wind field of

10   the lowest 14 vertical levels (out of 157 levels in total); these 14 levels represent the altitudes from surface to about 500 m (Hersbach and Dee, 2016;Beirle et al., 2011). Over the study period, the prevailing wind is northwesterly, and the wind speed is small over land (Fig. A1). For both zonal and meridional wind speeds, the uncertainty in the average wind speed is set to be 10%, which is similar to the temporal standard deviation of the wind speed. We assess the model errors introduced by the uncertainties in the

15   wind field and effective diffusion coefficients by Monte Carlo simulations; the resulting relative uncertainty in the modeled NO$_2$ VCDs is about 5%.

Re-mapping of PHLET simulated NO$_2$ VCDs in accordance to satellite pixels is important. Given the size of OMI pixels, the OMI NO$_2$ data smooth to some extent the actual horizontal distribution of NO$_2$. To ensure consistent spatial sampling between PHLET and OMI data, for each day we project the PHLET

20   modelled NO$_2$ VCD data (in the original 0.05° × 0.05° grid space) to the satellite pixel space to mimic how OMI "sees" the ground, remove the pixels with invalid OMI data, and then project the model data back to the 0.05° × 0.05° grid. The last two procedures are the same as done for OMI data. The whole process of grid conversion is done through the SCM approach (Appendix B). Although PHLET simulates summer average NO$_2$ VCDs (rather than daily values), we repeat the grid conversion process for as many

25   days as there are valid OMI data.

The model error σ_m is set to be the sum in quadrature of errors contributed by the steady state assumption (15%), the time averaging over multiple years and months (10%), the assumption of horizontally constant vertical shape of NO$_2$ (15%), the treatment of background NO$_2$ concentration (5%), and the error in the wind data and the calculation of effective diffusion coefficients (5%).

5   2.4.   PHLET-A: The adjoint model of PHLET

We construct the PHLET-A adjoint model to obtain an optimized horizontal distribution of the local net source term (L in Eq. (1)) under the given OMI NO$_2$ VCDs, wind field, and other parameters. PHLET-A accounts for the complex nonlinear effects of 2-dimensional transport and loss processes. We define a scalar cost function (Eq. (3)) to quantify the difference between OMI NO$_2$ VCDs and PHLET simulated

10   NO$_2$ VCDs.

$$J = (\mathbf{C}^{OMI} - \mathbf{C}^{PHLET})^T \mathbf{S}_o^{-1} (\mathbf{C}^{OMI} - \mathbf{C}^{PHLET}) \tag{3}$$

Because PHLET does not require a priori knowledge about the local net source, the cost function does not include the a priori term either. The vector $\mathbf{C}$ denotes gridded NO$_2$ VCDs. $\mathbf{S}_o$ denotes the observational error covariance matrix, as derived in Appendix C.

15   Applying Lagrange identity and integrating by parts, we derive PHLET-A and its initial and lateral boundary conditions (Marchuk, 1994;Sandu et al., 2005;Martien et al., 2006;Hakami et al., 2007):

$$\frac{\partial \lambda(x,y)}{\partial t} + \nabla(\mathbf{V}(x,y) \cdot \lambda(x,y)) + \nabla \cdot (\mathbf{K}(x,y) \cdot \nabla \lambda(x,y)) = 0 \tag{4}$$

$$\lambda(x,y)|_{t=T} = \frac{\delta J}{\delta C(x,y)}\Big|_{t=T} \tag{5}$$

$$\lambda(x,y)|_{boundary} = 0 \tag{6}$$

Here, T stands for the time when the domain-wide NO_2 VCDs come to equilibrium, i.e., the start time of the adjoint simulation. By discrete adjoint sensitivity analysis, the gradient of cost function to emission is obtained:

$$\frac{\delta J}{\delta L_{i,j}} = r \cdot \Delta x \cdot \Delta y \cdot \Delta t \cdot \sum_k \lambda_{i,j,k} \qquad (7)$$

5   where the indices i, j, k denote zonal, meridional, and time, respectively. The gradient is then used in an iterative optimization algorithm (Appendix C) to minimize the cost function J, i.e., to minimize the weighted difference between model simulated and OMI NO_2. The iteration stops when J is reduced from an initial value of 7937 to a stabilized value below 390 (Fig. C1).

2.5.   Deriving emission and loss from the local net source term

10   The optimized local net source term combines the contributions of emission (anthropogenic + soil) and loss (chemical loss + deposition). We further separate emission from loss by assuming a fixed formula within our small study domain for the nonlinear relationship between lifetime and VCD of NO_2.

In the summertime daytime, the dominant sink of NO_x is reactions with the radicals to produce nitric acid and organic nitrogen species. The NO_x chemistry quickly reaches a steady state under high solar radiation

15   and air temperature in the early afternoon (Murphy et al., 2006;Valin et al., 2013) when OMI passes over the YRD. The chemical lifetime of NO_x depends on the concentrations of NO_x and non-methane volatile organic compounds (NMVOC), radiation, temperature, and other factors. Within our small study domain, we assume all factors except NO_x concentrations to be spatially homogeneous. As such, the chemical lifetime of NO_x at steady state is a sole function of NO_x concentration (and thus NO_2 VCD, given the

20   constant NO_2/NO_x ratio). Appendix D shows in detail how to deduce the chemical lifetime of NO_x from NO_2 VCD, to account for the effect of dry deposition, to separate emission and lifetime from the local net source term, and to quantify the errors involved using Monte Carlo simulations.

2.6. Uncertainty estimate for inverted emissions

For a particular grid cell, our inverted emission is affected by the error involved in the inversion of L (embedded in satellite data and model simulations) and the error in the separation of emission and lifetime from L. The satellite data error σ_s is analyzed in Sect. 2.2. The model induced error σ_m is analyzed in

5   Sect. 2.3. The error of L, σ_L, is connected with σ_s and σ_m through the adjoint simulation (Sect. 2.4), and is given by Hessian of the cost function J (Appendix C).

The error involved in the separation of emission and lifetime, σ_f , is contributed by the assumption on the NO_2/NO_x ratio r (Sect. 2.3), the simplified treatment of deposition and chemical processes of NO_x (Appendix D), and the assumed relationship between lifetime and VCD (Appendix D). σ_f is estimated by

10   500 times of data fitting of L at individual grid cells (Appendix D).

Thus, the error in inverted emission, σ_e, is equal to the sum in quadrature of σ_L and σ_f, i.e., $\sigma_e = \sqrt{\sigma_L{}^2 + \sigma_f{}^2}$. The error in the inverted lifetime is derived in a similar way.

Appendix E further presents an OSSE-like test to estimate the reliability of our emission inversion method, by examining to what extent the method can invert the emissions used in a nested GEOS-Chem simulation

15   (0.3125° long. × 0.25° lat.) for summer 2014. Overall, the inverted emission data capture the mean value and spatial variability of the GEOS-Chem emissions, with a linear regression slope of 0.98, a correlation of 0.93, and a mean bias of 0.25 kg km^{-2} h^{-1} (16%). This test suggests the high reliability of our inversion method.

3. High-resolution spatial distributions of NO_2 VCDs, local net sources, and lifetimes over the
20   YRD

Figure 1a shows the number of days with valid OMI data in summer 2012–2016 over the YRD area on the 0.05°×0.05° grid. The number of days varies from about 14 to 113 (56 on average). There is a "band" pattern in the spatial distribution, due to the difference in the number of satellite orbits covering each grid

cell (not shown). This band pattern is not obvious in the distribution of OMI NO_2 VCDs (Fig. 1b),
suggesting that the temporally averaged VCD values are less sensitive to the number of days (14 or more)
used for temporal averaging. There are fewer valid data in severely polluted locations. The effect of
sampling size on the uncertainty in OMI NO_2 is accounted for in our study (Sect. 2.2).

5   Figure 1b shows the gridded horizontal distribution of OMI NO_2 VCDs. The background value has not
been removed. NO_2 VCDs are high over the major urban centers along the Yangtze River and the coastal
line, especially Shanghai, Nanjing (Capital of Jiangsu Province), Hangzhou (Capital of Zhejiang
Province), and the Ningbo-Zhoushan area (with intensive maritime shipping activities). The maximum
VCD value exceeds 14×10^{15} molecules cm^{-2} in north Shanghai. NO_2 VCDs are larger than 1×10^{15}

10 molecules cm^{-2} at all grid cells, reflecting the influence of local anthropogenic sources and/or pollution
transported from nearby cities (Cui et al., 2016). NO_2 VCDs are lower than 3×10^{15} molecules cm^{-2}
along the boundaries of our study domain.

Across the grid cells, the absolute errors in OMI NO_2 VCDs are about 0.58– 3.4×10^{15} molecules cm^{-2}
(Fig. 2a), and the relative errors are about 24%–54% (Fig. 2b). In general, the grid cells with larger NO_2

15 VCDs have larger absolute errors but smaller relative errors. Over the eastern sea and the southwestern
corner of the domain, NO_2 VCDs are relatively small (Fig. 1b), thus their absolute errors are small (Fig.
2a), but their relative errors are large (Fig. 2b).

Figure 1c shows the spatial distribution of the inverted local net source L (emission – loss). The values of
L are greatest (8.9 kg km^{-2} h^{-1}) over the major urban areas with high NO_2 VCDs, and are low (<-3.2 kg

20 km^{-2} h^{-1}) in many areas with low NO_2 loadings. However, the values of L are the lowest ((-7.2)–(-3.2) kg
km^{-2} h^{-1}) at places in the urban-rural fringe zones with NO_2 hotspots nearby; this feature reflects that NO_2
is transported from the urban centers and destroyed in the fringe zones. The absolute errors of L (Fig. 2c)
are highly consistent with those of the NO_2 VCD (Fig. 2a), with a spatial correlation of 0.95. The absolute
errors of L are notable in the urban-rural fringe zones, because the deviation of L at these areas is very

25 sensitive to errors in the assumed transport and loss process.

Figure 1d shows the inverted lifetime of NO_2 on the 0.05°×0.05° grid. The lifetime ranges from 1.4 to 3.6 h across the study domain. The lifetime is below 1.5 h at grid cells with NO_2 VCDs of about 2×10^{15} molecules cm^{-2}, increasing to 2.2 h at grid cells with the lowest VCDs ($< 2 \times 10^{15}$ molecules cm^{-2}), and exceeding 3 h at many polluted grid cells, i.e., the urban centers. The mean lifetime over the study domain is about 2.0 h. Appendix D further shows the chemical lifetime to be 1.7–6.6 h and the deposition lifetime to be constantly at 7.8 h across all grid cells. The nonlinear dependence of lifetime on VCD is expected from our inversion method (Appendix D). The absolute uncertainty in the inverted lifetime is greater than 1.0 h at the NO_2 hotspot locations, between 0.75 and 1.0 h over the eastern sea and the southwest of the study domain, and about 0.7 h at many other locations (Fig. 2d).

4. High-resolution spatial distribution of NO_x emissions over the YRD

4.1. Spatial distribution of inverted emissions

Figure 1e shows the inverted horizontal distribution of summer 2012–2016 average NO_x emissions on the 0.05°×0.05° grid. The inverted emissions include the contributions of anthropogenic and soil sources. As discussed in Sect. 4.3, soil emissions contribute little (0.9%) to the total emission over the study domain.

Figure 1e shows that NO_x emissions vary from 0 to 12.0 kg km^{-2} h^{-1} across the grid cells. The highest emission value occurs in north Shanghai, close to Wusongkou where Shanghai Port is located, which has become the largest container terminal all over the world in 2010 (Fu et al., 2012). High emission values also occur places along the Yangtze River and the coastal line. Along the Yangtze River, the highest emission value occurs in Nanjing City. Along the coastal line, there is an emission hotspot in the Ningbo-Zhoushan area. The general spatial distribution of inverted NO_x emissions is consistent with that of OMI NO_2 VCDs (correlation = 0.82), reflecting the short lifetime of NO_x and thus a modest effect of horizontal transport. Nonetheless, emissions are much more concentrated at a few sparse locations than NO_2 VCDs are, and many locations near the emission hotspots have very low emissions but relatively large NO_2 VCDs, suggesting that the effect of horizontal transport cannot be ignored at such a high resolution.

Figure 2e shows the absolute uncertainties of NO_x emissions at individual grid cells. The emission uncertainty varies from 0.1 to 1.4 kg km^{-2} h^{-1} across all grid cells. The largest uncertainty occurs in north Shanghai, corresponding to the highest VCD (Fig. 1b) and emission (Fig. 1e) values. The spatial pattern of emission uncertainties (Fig. 2e) is closer to the pattern of NO_2 VCDs (Fig. 1b, correlation = 0.91) than

5   to the pattern of emissions (Fig. 1e, correlation = 0.75). The emission uncertainties are spatially more homogeneous than VCDs and emissions are, because the emission uncertainties can be high at locations with low VCDs and emissions. The spatial pattern of emission uncertainties is consistent with that of L uncertainties (Fig. 2c, correlation = 0.85).

The scatter plot in Fig. 2f shows the emission uncertainty as a function of emission at each grid cell.

10   Although the emission uncertainty tends to be higher when the inverted emission value increases, the relationship is highly nonlinear, and there is large data spread. The data spread reaches its maximum value when emissions are the lowest, declines with increasing emissions, and is relatively small when emissions exceed 3 kg km^{-2} h^{-1}. The data points in Fig. 2f are further colored to indicate the VCDs at individual grid cells. It is clear that the grid cells with low emissions but high emission uncertainties have large NO_2

15   VCDs. These grid cells are located in the urban-rural fringe zones with significant NO_2 pollution transported from nearby urban centers.

4.2.   Comparing our inverted emissions with nighttime light, population density, and road network

This section compares our inverted NO_x emission dataset with several spatial proxies widely used in bottom-up inventories, including nighttime light brightness ($0.5' \times 0.5'$, Fig. 3a), population density

20   ($0.1° \times 0.1°$, Fig. 3b), and road network (Fig. 3c). These proxies broadly represent the intensity of human activities and are highly related to NO_x emissions (Geng et al., 2017).

Figure 3a shows the spatial distribution of nighttime light brightness in 2012. The data are taken from Version 4 DMSP-OLS Nighttime Lights Time Series (https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html; last access: 2018/08/19). The

25   brightness is represented digitally from 0 to 63 bits. The nighttime light reflects the intensity of household

activity, commercial activity, and resource consumption (Elvidge et al., 2013). When regridded to $0.05° \times 0.05°$, the spatial correlation between nighttime light brightness and NO_x emissions is about 0.70, i.e., the nighttime light brightness can explain 49% of the spatial variability in NO_x emissions in the YRD area.

5   Figure 3b shows the population density data, which are taken from the Gridded Population of the World v4 (GPWv4) (http://sedac.ciesin.columbia.edu/data/collection/gpw-v4/sets/browse; last access: 2018/08/19) (Center for International Earth Science Information Network - CIESIN - Columbia University, 2016). This dataset provides population density data for every five years (2000, 2005, 2010, 2015, etc.). Data in 2012, 2013 and 2014 are estimated by fitting a natural spline to the 2000, 2005, 2010,
10 and 2015 values. Data in 2016 are estimated by fitting a natural spline to the 2005, 2010, 2015, and 2020 values. The population density varies greatly from the urban centers to the countryside. In north Shanghai, the population density exceeds 5×10^3 km^{-2}. The NO_x emission hotspots match the population hotspots, and the lowest-emission locations have little population. When regridded to $0.05° \times 0.05°$, the spatial correlation between population densities and NO_x emissions is 0.52.

15 Figure 3c shows our NO_x emissions overlaid with the OpenStreetMap road network data (http://download.geofabrik.de; last access: 2018/6/27). The network includes both highways and local roads. In the southern areas (between 29°N and 31°N), the spatial distribution of NO_x emissions largely coincides with the road network. The spatial coincidence is less obvious in the north because of the influence of non-mobile sources. NO_x emissions are notable along the three major national highways
20 connecting Jinhua City (one of the largest hubs of light industry products in China), Hangzhou City, and Ningbo City. NO_x emissions are also identifiable along the national highway from Hangzhou City to Huangshan City. Pairs of NO_x emissions and traffic hubs are located to the west of the Taihu Lake and in the urban centers. These results suggest the capability of our inverted emission dataset in capturing the contribution of traffic sources.

4.3. Comparing our inverted emission dataset to the MEIC inventory

This section compares our inverted emission data to the widely used MEIC anthropogenic bottom-up inventory over summer 2012–2016 (www.meicmodel.org; last access: 2018/7/2). The MEIC data are available at the $0.25° \times 0.25°$ resolution. Because our emission data include the slight contribution of soil sources, we use the nested GEOS-Chem to calculate soil emissions (Fig. 3d), and then subtract these soil emissions from our data. Figure 3ows the resulting "anthropogenic" portion of our inverted emissions. Summed over the study domain, the soil sources contribute 0.9% of our inverted emissions.

Figures 3e and 3f show that compared to MEIC, our high-resolution anthropogenic emission dataset provides much more detailed spatial information. Our dataset identifies the emission hotspots and their contrast with nearby low-emission areas (e.g., in the urban-rural fringe zones) better than MEIC does. The contribution of mobile sources along the road network is clearer in our dataset. Our emission data contain sources over the nearby sea (i.e., from shipping), along the coastal line, and in the southwest of the domain, which are not included in MEIC. In Hangzhou City, our dataset suggests the highest emissions to be to the south of Qiantang River (marked in purple line in Fig. 1b), whereas the MEIC inventory suggests north. The total emission in MEIC is smaller than ours by 28.3% over the entire study domain and by 10.4% over land.

Figure 4 compares emissions in several cities between our and MEIC inventory. A total of 14 cities within the domain are selected, and for each city the NO_x emissions are summed over the grid cells within the municipal administrative boundaries. The MEIC inventory is re-gridded to $0.05° \times 0.05°$ for this purpose. Among the cities, emission values differ from -51.4% to +61.3% between the two datasets. In Shanghai, Suzhou and Wuxi, our dataset is lower than MEIC by 27.1%, 43.7%, and 51.4%, respectively. MEIC uses the industrial GDP as a spatial proxy to attribute provincial-level industrial emissions to individual counties and grid cells (Geng et al., 2017). This may overestimate the emissions in Suzhou and Wuxi, since much of the industrial GDP in the two cities is contributed by high-tech industries with low emissions.

Figure 4 also shows that in Zhoushan, Huai'an and Yancheng, our dataset is higher than MEIC by 61.3%, 55.5%, and 55.1%, respectively. Zhoushan (around 30°N, 122°E) has many isles and marine ports, as identifiable on the nighttime light map (Fig. 3a). The marine ports in the Ningbo-Zhoushan area contribute about 10% of the total shipping emissions in China (Endresen et al., 2003;Fu et al., 2017). Our dataset accounts for emissions from marine shipping and ports, whereas MEIC does not.

Section 4.2 shows the spatial correlation at the 0.05° × 0.05° resolution between our inverted NO_x emissions and the proxies to be 0.70 for nighttime light and 0.52 for population density. When regridded to 0.25° × 0.25°, the correlation becomes higher: 0.79 for nighttime light and 0.63 for population density. As the spatial resolution gets finer, the chance that NO_x emissions are collocated with population or nighttime light becomes smaller (Zheng et al., 2017), because of the influences of NO_x-emitting factories, power plants, and mobile sources. By comparison, the correlation between the MEIC inventory and these proxies on the 0.25° × 0.25° grid is 0.83 for nighttime light and 0.71 for population density, higher than for our emission dataset. This difference is in part because our dataset better accounts for the influences of land mobile and offshore shipping sources which are spatially not tied closely to nighttime light and population at this resolution.

5. Concluding remarks

This study presents a satellite-based NO_x emission inversion method for urban and surrounding areas at very high horizontal resolution. As a demonstration, the method is applied to the YRD area at the 0.05° × 0.05° resolution in summer 2012–2016, based on the POMINO NO_2 product. We construct a simplified, computationally efficient 2-D lifetime-emission-transport model (PHLET) and its adjoint model (PHLET-A) to, together with other procedures, facilitate the emission inversion. The reliability of our inversion method is supported by 1) a rigorous step-by-step derivation of models, assumptions, and parameters used, 2) a comprehensive uncertainty analysis, and 3) an OSSE-like test with GEOS-Chem simulated NO_2 data. Our inverted emission dataset in the YRD area on the 0.05° × 0.05° grid shows fine-scale spatial information that is tied to nighttime light, population density, road network, and maritime

shipping. Our dataset reveals many fine-scale spatial characteristics not well represented or not included in the widely used MEIC inventory.

Our emission inversion method is useful for understanding how human activities have altered the atmospheric environment at fine resolutions. Many crucial human activities, such as urbanization, are conducted at very fine spatial scales. How the resulting emissions affect air quality, public health, and geo-health are still poorly understood due to lack of high-resolution emission data. This problem is particularly severe in the developing countries, because of their rapid paces of urbanization and great inadequacies in emission-related information such as economic statistics and emission factors. This poses a grand challenge for emission control and environmental management. Thus, our inversion method and resulting emission data offer useful independent high-resolution information to monitor the fine-scale emission sources, to improve the bottom-up inventory, to model the urban pollution chemistry and the effect of urbanization, and to conduct spatially targeted emission control.

Our emission inversion method and models have a few important features enabling their global applications. The PHLET model offers computationally efficient simulations of the NO_x chemistry, deposition, and transport. At a low computational cost, our inversion method is able to account for the nonlinear relationship between NO_x concentration, chemical loss, deposition, and transport. With the advent of TROPOMI and other satellite sensors with unprecedented spatial resolutions, our inversion approach can be applied to these measurements for continuous inference of emissions at finer and finer resolutions. As a final point, PHLET and PHLET-A are written in the Python language, which can be readily used with low financial costs.

Data availability

Observational data are obtained from individual sources (see links in the text and acknowledgments). Model results are available upon request. Model codes are available on a collaborative basis.

Author contributions

JL conceived the research. HK, RZ and JL designed the research. HK and RZ performed the data processing, model development, and simulations. ML, HW, LC, RN and JW contributed to data processing, model simulations, and data analyses. QZ provided MEIC data. HK and JL analyzed the

5 results and wrote the paper with input from all authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (41775115) and the 973

10 program (2014CB441303).

Appendix A. Solving the diffusion process

The diffusion term can be simplified as follows:

$$\nabla \cdot \left(\mathbf{K}(x,y) \cdot \nabla C(x,y) \right) = \frac{\partial}{\partial x}\left(K_x \cdot \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y}\left(K_y \cdot \frac{\partial C}{\partial y} \right) \tag{A1}$$

K_x and K_y are the diffusion coefficients in the zonal and meridional directions, respectively. Given these

15 coefficients, we solve the equation with the algorithm described in (Brasseur and Jacob, 2017).

We derive the diffusion coefficients based on a random walk assumption (Schirmacher, 2015):

$$K_{x \, or \, y} = \frac{1}{2} \overline{V_{x \, or \, y}'}^2 \, t_0, \tag{A4}$$

$\overline{V_{x\,or\,y}}'$ is the deviation of wind speed in the zonal or meridional direction. t_0 is 3 hours, the sampling interval of ERA5 wind data. Figure A1 shows the time averaged wind vector and the distribution of K_x and K_y. The relative uncertainty in wind speed is assumed to be 10%, close to the temporal standard deviation of wind speed. The uncertainties of K_x and K_y are set to be 20%, about twice of the relative

5   uncertainty in wind speed. The calculated K_x ranges from $30397\ m^2 s^{-2}$ over land to $203783\ m^2 s^{-2}$ over sea. The K_y ranges from $25811\ m^2 s^{-2}$ over land to $297053\ m^2 s^{-2}$ over sea. These diffusion coefficients tend to be slightly underestimated, because the variabilities of wind speed at higher frequencies (than 3-hourly) are not accounted for. This means that PHLET may underestimate the horizontal transport slightly.

10   **Appendix B. Satellite Conversion Matrix to account for the smoothing effect of satellite pixels**

The SCM is essentially a tool to preform quick conversion between grids, regular or not. In the YRD area, there are $100\ \times\ 100\ =\ 10000$ grid cells on the $0.05° \times 0.05°$ grid. We use the SCM (**A** matrix: [10000, 10000]) to convert from its original grid (**X** vector: [10000, 1]) to the final grid (**Y** vector: [10000, 1]), i.e., $Y = AX$. The 10000 elements in one specific row of **A** represent the weights of the 10000 elements of **X**

15   to an element in **Y**. Apparently, **A** is a sparse matrix. The following description shows how **A** is constructed.

First, the VCDs specific to satellite pixels are reconstructed from the model grid cells. Each model grid cell (MGC) is divided into $10\ \times\ 10\ =\ 100$ finer grid cells (FGCs), each having the same area. Suppose the number of MGCs fully or partially covered by a given pixel p is N_c, and the number of FGCs in a

20   given MGC i covered by p is g_i^p, then the total number of FGCs covered by p is:

$$G^p = \sum_{i=1}^{N_c} g_i^p \tag{B1}$$

Thus, the average VCD for the pixel p can be reconstructed as follows:

$$VCD^p = \sum_{i=1}^{N_c} \frac{g_i^p}{G^p} \cdot VCD_i \qquad (B2)$$

Equation (B2) essentially means how a satellite pixel smooths the VCD. The blue portion of Fig. B1 denotes the projection from MGC i to pixel p.

The next step represents how the oversampling approach is applied to satellite-smoothed VCD data.

5   Suppose the number of satellite pixels fully or partially covering a MGC j is Np, then the total number of FGCs being part of the intersection of the Np pixels and MGC j is:

$$G_j = \sum_{p=1}^{N_p} g_j^p \qquad (B3)$$

Finally, the average VCD for the MGC j converted from the Np pixels is:

$$VCD'_j = \sum_{p=1}^{N_p} \frac{g_j^p}{G_j} \cdot VCD^p \qquad (B4)$$

10   The pink portion of Fig. B1 denotes the projection from pixel p to MGC j. Thus, the element of SCM converting from MGC i to MGC j can be derived as follows:

$$A_{j,i} = \sum_{p=1}^{N_p} \frac{g_j^p}{G_j} \cdot \frac{g_i^p}{G^p} \qquad (B5)$$

Appendix C. Solving the observation error covariance matrix and the adjoint model

The observation error covariance matrix ($\mathbf{S_o}$) consists of a satellite data error covariance matrix ($\mathbf{S_s}$) and

15   a PHLET model error covariance matrix ($\mathbf{S_m}$):

$$\mathbf{S_o} = \mathbf{S_s} + \mathbf{S_m} \qquad (C1)$$

Both $\mathbf{S_s}$ and $\mathbf{S_m}$ are assumed to be diagonal, with the diagonal elements set to be σ_s^2 and σ_m^2, respectively.

As shown in Eq. (7), PHLET-A represents the sensitivity of cost function (J) to local net source (L) (Marchuk, 1994;Sandu et al., 2005). Given a starting point of L , we can derive a search direction by the Cautious Broyden-Fletcher-Goldfarb-Shanno (CBFGS) method (Li and Fukushima, 2001;Bousserez et al., 2015). Then, by practicing backtracking line search based on the Armijo–Goldstein condition (Armijo, 1966), we can evaluate an optimized L. The uncertainty of L is given by the Hessian of the cost function, which is approximated by the CBFGS method (Brasseur and Jacob, 2017):

$$S = 2 \cdot (\nabla_L^2 J)^{-1} \tag{C2}$$

Appendix D. Deriving NO₂ lifetime from VCD

We assume a steady state of radicals (HO$_x$), where the production rate of HO$_x$ is equal to the loss rate through three types of termination reactions: between the hydroxyl radical (OH) and NO₂, between NO and peroxyl radicals to form organic nitrates, and between peroxyl radicals (Murphy et al., 2006;Valin et al., 2011):

$$P(HO_x) = k_1 C_{OH} C_{NO_2} + \alpha k_{2eff} \frac{k_1 C_{NMVOC} C_{OH}}{k_{2eff} C_{NO}} C_{NO} + 6 k_{3eff} \left(\frac{k_1 C_{NMVOC} C_{OH}}{k_{2eff} C_{NO}}\right)^2 \tag{D1}$$

Here $P(HO_x)$ is the production rate, and the right hand side of Eq. (D1) is the loss rate. C_{NO_2} and C_{OH} denote the concentrations of NO₂ and OH, respectively. Since the conversion between the peroxyl radicals (HO₂ + RO₂) and OH is in steady state, the term $\frac{k_1 C_{NMVOC} C_{OH}}{k_{2eff} C_{NO}}$ expresses the "effective" total concentration of peroxyl radicals in terms of the concentrations of NMVOC, OH and NO. Assuming $P(HO_x)$, C_{NMVOC} and all reaction constants to be constant (Valin et al., 2011), and given that $C_{NO} = C_{NO_2} \cdot \frac{1-r}{r}$, Eq. (D1) can be simplified as Eq. (D2):

$$a C_{OH} C_{NO_2} + b C_{OH} + c \left(\frac{C_{OH}}{C_{NO_2}}\right)^2 = 1 \tag{D2}$$

Here a, b, c are the coefficients. Because the chemical lifetime of NO₂ is determined by C_{OH} $(\tau_c = \frac{1}{k_1 C_{OH}})$, we can deduce the relationship between C_{NO_2} and τ_c:

$$\frac{a\prime}{\tau_c} + \frac{b\prime}{\tau_c \cdot C_{NO_2}} + c\prime(\frac{1}{\tau_c \cdot C_{NO_2}})^2 = -1 \tag{D3}$$

NO₂ is lost primarily through reaction with OH and secondarily through dry deposit $(\frac{C_{NO_2}}{\tau_d})$, thus its

5 lifetime (τ) is also determined by these two loss processes. Therefore,

$$\tau_c = \frac{1}{\frac{1}{\tau} - \frac{1}{\tau_d}} \tag{D4}$$

In the areas of low emissions, the emission term can be neglected in Eq. 2, thus the local net source $L = -\frac{C_{NO_2}}{r \cdot \tau}$. Therefore Eq. D3 becomes Eq. D5, which connects L and C_{NO_2}.

$$a\prime \frac{r(L+kC_{NO_2})}{C_{NO_2}} + b\prime \frac{r(L+kC_{NO_2})}{C_{NO_2}^2} + c\prime(\frac{r(L+kC_{NO_2})}{C_{NO_2}^3})^2 = 1 \tag{D5}$$

10 where $k = \frac{1}{\tau_d}$. We determine the coefficients $a\prime, b\prime, c\prime$, and k in Eq. D5 by conducing nonlinear fitting of OMI NO₂ VCD data and L values in the low emission areas (see below). This procedure establishes the nonlinear relationship between τ and VCD, which is then applied to the entire study domain.

The low-emission areas have small values of VCD and large negative values of L. Figure D1a shows a scatter plot for the inverted local net source L and VCD at each individual grid cell of the study domain.

15 The data scatter reflects the combined effect of emission, loss, and horizontal transport. We then fit the quantiles of L where the VCD is relatively low ($< 5 \times 10^{15}$ molecules cm⁻²) into Eq. 6 through a nonlinear quantile fitter based on Tensor flow. Using the quantile fitting also means that the low-emission grid cells do not need to be explicitly identified prior to the fitting. The quantile fitting gives L as a function of VCD (when emissions are neglected), through which the relationship between lifetime and VCD is derived. We

conduct the fitting by 500 times, each by linearly changing the assumed percentile threshold of *L* from 0.01% to 10%, to determine the fitted median value (red line in Fig. D1a) and uncertainty (gray shaded areas, 95% CI). The uncertainty is caused by the assumption on the NO_2/NO_x ratio r, the simplification of the relationship between lifetime and VCD, and possible misjudgment of low-emission areas.

5   The orange line in Fig. D1b presents the relationship between NO_2 VCD and chemical lifetime (τ_c) derived based on the mean value of the fitting. The value of τ_c varies from 1.7 to 6.6 h with an average of 2.8 h. The lifetime declines rapidly with increasing VCD from 0 to 2×10^{15} molecules cm^{-2}, and then grows gradually with increasing VCD. This result is consistent with Valin et al. (2011). By comparison, the value of τ_d is 7.8 h and is spatially homogeneous under the assumption here. The total lifetime (τ)

10   varies from 1.4 to 3.6 h (Fig. D1b, blue line) across the study domain.

Appendix E. Testing our emission inversion method using GEOS-Chem simulated NO₂ data

We apply our emission inversion method to the NO_2 VCD data simulated by the nested GEOS-Chem CTM. Specifically, we use GEOS-Chem v9-02 (Liu et al., 2018a;Ni et al., 2018;Yan et al., 2016) to simulate the NO_2 VCDs in the early afternoon (around the overpass time of OMI) in summer 2014 on the

15   0.3125° × 0.25° grid. The simulated NO_2 data are shown in Fig. E1a. Next, we convert the GEOS-Chem NO_2 VCDs into the 0.05° × 0.05° grid, and parameterize PHLET with the wind field adopted by GEOS-Chem, following the procedures in Sect. 2.3. Then we use PHLET, PHLET-A, and the lifetime-emission separation method to estimate the NO_x emissions. Finally, we compare the inverted emissions (re-mapped to the 0.25° × 0.3125° grid) to those used in GEOS-Chem.

20   Figure E1c and E1d shows the horizontal distributions of our inverted "anthropogenic" emissions and emission uncertainties, respectively. The contribution of soil emissions (as simulated by GEOS-Chem) is subtracted from the inverted dataset. Figure E1e shows the differences between the inverted anthropogenic emissions and the GEOS-Chem anthropogenic emissions. The emission difference at each grid cell varies from −4.5 to 5.8 kg $km^{-2}\,h^{-1}$, which is attributed to the limitation of our inversion method.

The domain average difference is 0.25 kg km^{-2} h^{-1}, or 16% (with respect to GEOS-Chem emission value). The scatter plot in Fig. E1f suggests excellent consistency between the inverted and the GEOS-Chem emissions, with a linear regression slope of 0.98 and correlation of 0.93. The inverted emissions tend to be overestimated in the low-emission areas, but the differences are within the uncertainties of the inverted
5   emissions (shaded area).

References

NBSC (National Bureau of Statistics of China) STATISTICAL COMMUNIQUÉ OF THE PEOPLE'S REPUBLIC OF CHINA ON THE 2017 NATIONAL ECONOMIC AND SOCIAL DEVELOPMENT,

10   NBSC (National Bureau of Statistics of China) STATISTICAL COMMUNIQUÉ OF THE PEOPLE'S REPUBLIC OF CHINA ON THE 2013 NATIONAL ECONOMIC AND SOCIAL DEVELOPMENT,

NBSC (National Bureau of Statistics of China) STATISTICAL COMMUNIQUÉ OF THE PEOPLE'S REPUBLIC OF CHINA ON THE 2009 NATIONAL ECONOMIC AND SOCIAL DEVELOPMENT,

NBSC (National Bureau of Statistics of China) China Statistical Yearbook 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017,

Armijo, L. J. P. J. o. M.: Armijo, L.: Minimization of functions having Lipschitz continuous first partial derivatives. Pac. J. Math. 16, 1-3,
15   16, 1966.

Barnes, I., and Rudziński, K. J.: Disposal of Dangerous Chemicals in Urban Areas and Mega Cities, Springer Netherlands, 2013.

Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G., and Wagner, T.: Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space, Science, 333, 1737-1739, doi:10.1126/science.1207824, 2011.

Beirle, S., Sihler, H., and Wagner, T.: Mapping NOx Sources on High Spatial Resolution by Combined Measurements from OMI and
20   GOME-2 Narrow Swath Mode, Atmos, 2015,

Boersma, K., Vinken, G., and Tournadre, J.: Ships going slow in reducing their NOx emissions: Changes in 2005-2012 ship exhaust inferred from satellite measurements over Europe, 2015.

Boersma, K. F., Eskes, H. J., Dirksen, R. J., and R., J. v. d. A. J. A. M. T.: An improved tropospheric NO2 column retrieval algorithm for the Ozone Monitoring Instrument, 4, 2329-2388, 2011.

25   Bousserez, N., Henze, D. K., Perkins, A., Bowman, K. W., Lee, M., Liu, J., Deng, F., and Jones, D. B. A.: Improved analysis-error covariance matrix for high-dimensional variational inversions: application to source estimation using a 3D atmospheric transport model, Quarterly Journal of the Royal Meteorological Society, 141, 1906-1921, doi:10.1002/qj.2495, 2015.

Brasseur, G. P., and Jacob, D. J.: Modeling of atmospheric chemistry, 2017.

Cui, Y. Z., Lin, J. T., Song, C., Liu, M. Y., Yan, Y. Y., Xu, Y., Huang, B. J. A. C., and Physics: Rapid growth in nitrogen dioxide pollution
30   over Western China, 2005-2013, 15, 34913-34948, 2016.

de Graaf, M., Sihler, H., Tilstra, L. G., and Stammes, P.: How big is an OMI pixel?, Atmos. Meas. Tech., 9, 3607-3618, doi:10.5194/amt-9-3607-2016, 2016.

Ding, J., Johannes, V. D. A., Ronald, Mijling, B., and Felicitas Levelt, P. J. A. M. T.: Space-based NOx emission estimates over remote regions improved in DECSO, 10, 1-21, 2017a.

Ding, J., Miyazaki, K., Johannes, V. D. A., Ronald, Mijling, B., Kurokawa, J., Cho, S. Y., Janssensmaenhout, G., Zhang, Q., Liu, F., Felicitas Levelt, P. J. A. C., and Physics: Intercomparison of NOx emission inventories over East Asia, 17, 1-35, 2017b.

Elvidge, C., Hsu, F. C., Baugh, K. E., and Ghosh, T.: National Trends in Satellite Observed Lighting: 1992-2012, in: lobal Urban Monitoring and Assessment Through Earth Observation (In Press), CRC Press, 2013.

Endresen, Ø., Sørgård, E., Sundet, J. K., Dalsøren, S. B., Isaksen, I. S. A., Berglen, T. F., and Gravir, G. J. J. o. G. R. A.: Emission from international sea transportation and environmental impact, 108, -, 2003.

Eskes, H. J., Velthoven, P. F. J. V., Valks, P. J. M., and Kelder, H. M.: Assimilation of GOME total-ozone satellite observations in a three-dimensional tracer-transport model, Quarterly Journal of the Royal Meteorological Society, 129, 1663-1681, doi:10.1256/qj.02.14, 2003.

Fioletov, V. E., McLinden, C. A., Krotkov, N., Moran, M. D., and Yang, K.: Estimation of SO_2 emissions using OMI retrievals, Geophysical Research Letters, 38, n/a-n/a, doi:10.1029/2011gl049402, 2011.

Fu, M., Liu, H., Jin, X., and He, K.: National- to port-level inventories of shipping emissions in China, Environmental Research Letters, 12, doi:10.1088/1748-9326/aa897a, 2017.

Fu, Q., Shen, Y., and Zhang, J.: On the ship pollutant emission inventory in Shanghai port, J. Saf. Environ., 12, 57–64, 2012.

Geng, G., Zhang, Q., Martin, R. V., Lin, J., Huo, H., Zheng, B., Wang, S., and He, K.: Impact of spatial proxies on the representation of bottom-up emission inventories: A satellite-based analysis, Atmospheric Chemistry and Physics, 17, 4131-4145, doi:10.5194/acp-17-4131-2017, 2017.

Gu, D., Wang, Y., Smeltzer, C., and Boersma, K. F.: Anthropogenic emissions of NO_x over China: Reconciling the difference of inverse modeling results using GOME-2 and OMI measurements, Journal of Geophysical Research: Atmospheres, 119, 7732-7740, doi:10.1002/2014jd021644, 2014.

Hakami, A., Henze, D. K., Seinfeld, J. H., Singh, K., Sandu, A., Kim, S., Byun, D., and Li, Q.: The Adjoint of CMAQ, Environ. Sci. Technol, 41, 7807-7817, 2007.

Hersbach, H., and Dee, D.: ERA5 reanalysis is in production, ECMWF Newsletter No. 147, 7, 2016.

Huang, M., Bowman, K. W., Carmichael, G. R., Chai, T., Pierce, R. B., Worden, J. R., Luo, M., Pollack, I. B., Ryerson, T. B., and Nowak, J. B. J. o. G. R.: Changes in Nitrogen Oxides Emissions in California during 2005-2010 Indicated from Top-down and Bottom-up Emission Estimates, 119, 12,928-912,952, 2014.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., and Wankmüller, R.: HTAP_v2.2: a mosaic of regional and global emission gridmaps for 2008 and 2010 to study hemispheric transport of air pollution, Atmospheric Chemistry & Physics, 15, 12867-12909, 2015.

Krotkov, N. A., McLinden, C. A., Li, C., Lamsal, L. N., Celarier, E. A., Marchenko, S. V., Swartz, W. H., Bucsela, E. J., Joiner, J., Duncan, B. N., Boersma, K. F., Veefkind, J. P., Levelt, P. F., Fioletov, V. E., Dickerson, R. R., He, H., Lu, Z., and Streets, D. G.: Aura OMI observations of regional SO_2 and NO_2 pollution changes from 2005 to 2015, Atmospheric Chemistry and Physics, 16, 4605-4629, doi:10.5194/acp-16-4605-2016, 2016.

Lamsal, L. N., Martin, R. V., Padmanabhan, A., van Donkelaar, A., Zhang, Q., Sioris, C. E., Chance, K., Kurosu, T. P., and Newchurch, M. J.: Application of satellite observations for timely updates to global anthropogenic NO(x) emission inventories, Geophysical Research Letters, 38, L05810, doi:10.1029/2010gl046476, 2011.

Li, D.-H., and Fukushima, M.: A modified BFGS method and its global convergence in nonconvex minimization, Journal of Computational and Applied Mathematics, 129, 15-35, 2001.

Lin, J.-T., Liu, M.-Y., Xin, J. Y., Boersma, K. F., Spurr, R., Martin, R., and Zhang, Q.: Influence of aerosols and surface reflectance on satellite NO2 retrieval: seasonal and spatial characteristics and implications for NOx emission constraints, Atmospheric Chemistry and Physics, 15, 11217-11241, doi:10.5194/acp-15-11217-2015, 2015.

Lin, J., Tong, D., Davis, S., Ni, R., Tan, X., Pan, D., Zhao, H., Lu, Z., Streets, D., Feng, T., Zhang, Q., Yan, Y., Hu, Y., Li, J., Liu, Z., Jiang, X., Geng, G., He, K., Huang, Y., and Guan, D.: Global climate forcing of aerosols embodied in international trade, Nature Geoscience, 9, 790-794, doi:10.1038/ngeo2798, 2016.

Lin, J. T., McElroy, M. B., and Boersma, K. F.: Constraint of anthropogenic NOx emissions in China from different sectors: a new methodology using multiple satellite retrievals, Atmos. Chem. Phys., 10, 63-78, doi:10.5194/acp-10-63-2010, 2010.

Lin, J. T.: Satellite constraint for emissions of nitrogen oxides from anthropogenic, lightning and soil sources over East China on a high-resolution grid, Atmos. Chem. Phys., 12, 2881-2898, doi:10.5194/acp-12-2881-2012, 2012.

Lin, J. T., Liu, Z., Zhang, Q., Liu, H., Mao, J., and Zhuang, G.: Modeling uncertainties for tropospheric nitrogen dioxide columns affecting satellite-based inverse modeling of nitrogen oxides emissions, Atmos. Chem. Phys., 12, 12255-12275, doi:10.5194/acp-12-12255-2012, 2012.

Lin, J. T., Martin, R. V., Boersma, K. F., Sneep, M., Stammes, P., Spurr, R., Wang, P., Roozendael, M. V., Clémer, K., Irie, H. J. A. C., and Physics, 3: Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: Effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide, 14, 1441-1461, 2014.

Liu, F., Beirle, S., Zhang, Q., Dörner, S., He, K., and Wagner, T.: NOx lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations, Atmospheric Chemistry and Physics, 16, 5283-5298, doi:10.5194/acp-16-5283-2016, 2016.

Liu, M.-Y., Lin, J.-T., Wang, Y.-C., Sun, Y., Zheng, B., Shao, J., Chen, L.-L., Zheng, Y., Chen, J., Fu, T. M., Yan, Y.-Y., Zhang, Q., and Wu, Z.: Spatiotemporal variability of NO2 and PM2.5 over Eastern China: observational and model analyses with a novel statistical method, Atmospheric Chemistry and Physics, 18, 12933-12952, doi:10.5194/acp-18-12933-2018, 2018a.

Liu, M., Lin, J., Boersma, K. F., Pinardi, G., Wang, Y., Chimot, J., Wagner, T., Xie, P., Eskes, H., Van Roozendael, M., Hendrick, F., Wang, P., and Yan, Y.: Improved aerosol correction for OMI tropospheric NO₂ retrieval over East Asia: constraint from CALIOP aerosol vertical profile, Atmospheric Measurement Techniques Discussions, 1-48, doi:10.5194/amt-2018-34, 2018b.

Marchuk, G. L.: Adjoint Equations and Analysis of Complex Systems, Russian Academy of Sciences, Moscow, Russia, 475 pp., 1994.

Martien, P. T., Harley, R. A., and Cacuci, D. G.: Adjoint Sensitivity Analysis for a Three-Dimensional Photochemical Model: Implementation and Method Comparison, Environ. Sci. Technol, 40, 2663-2670, 2006.

Martin, R. V., Jacob, D. J., Chance, K., Kurosu, T. P., Palmer, P. I., and Evans, M. J.: Global inventory of nitrogen oxide emissions constrained by space-based observations of NO2 columns, Journal of Geophysical Research: Atmospheres, 108, 4537, doi:10.1029/2003JD003453, 2003.

Mijling, B., and R., J. v. d. A. J. J. o. G. R. A.: Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale, 117, -, 2012.

Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NOₓ emission estimates derived from an assimilation of OMI tropospheric NO₂ columns, Atmospheric Chemistry and Physics, 12, 2263-2288, doi:10.5194/acp-12-2263-2012, 2012.

5   Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NOₓ emissions from multi-constituent satellite data assimilation, Atmospheric Chemistry and Physics Discussions, 1-48, doi:10.5194/acp-2016-529, 2016.

Murphy, J. G., Day, D. A., Cleary, P. A., Wooldridge, P. J., Millet, D. B., Goldstein, A. H., and Cohen, R. C.: The weekend effect within and downwind of Sacramento: Part 2. Observational evidence for chemical and dynamical contributions, Atmospheric Chemistry and Physics Discussions, 6, 11971-12019, doi:10.5194/acpd-6-11971-2006, 2006.

10   Ni, R.-J., Lin, J.-T., Yan, Y.-Y., and Lin, W.: Foreign and domestic contributions to springtime ozone over China, Atmospheric Chemistry and Physics, 18, 11447-11469, doi:10.5194/acp-18-11447-2018, 2018.

Qu, Z., Henze, D. K., Capps, S. L., Wang, Y., Xu, X., Wang, J., and Keller, M. J. J. o. G. R. A.: Monthly top-down NOx emissions for China (2005–2012): A hybrid inversion method and trend analysis, 122, 2017.

Sandu, A., Daescu, D. N., Carmichael, G. R., and Chai, T.: Adjoint sensitivity analysis of regional air quality models, Journal of
15   Computational Physics, 204, 222-252, doi:10.1016/j.jcp.2004.10.011, 2005.

Schere, K. L. J. A. E.: An evaluation of several numerical advection schemes, 17, 1897-1907, 1983.

Schirmacher, W.: Theory of Liquids and Other Disordered Media, Springer International Publishing, 2015.

Seinfeld, J. H., Pandis, S. N. J. A. C., and Physics: From air pollution to climate change, 429-443, 2006.

Stavrakou, T., Muller, J. F., Boersma, K. F., De Smedt, I., and van der A, R. J.: Assessing the distribution and growth rates of NOx emission
20   sources by inverting a 10-year record of NO2 satellite columns, Geophysical Research Letters, 35, L10801, doi:10.1029/2008gl033521, 2008.

Stavrakou, T., Müller, J. F., Boersma, K. F., van der A, R. J., Kurokawa, J., Ohara, T., and Zhang, Q.: Key chemical NOₓ sink uncertainties and how they influence top-down emissions of nitrogen oxides, Atmos. Chem. Phys., 13, 9057-9082, doi:10.5194/acp-13-9057-2013, 2013.

25   Sun, K., Zhu, L., Cady-Pereira, K., Chan Miller, C., Chance, K., Clarisse, L., Coheur, P.-F., Gonzalez Abad, G., Huang, G., Liu, X., Van Damme, M., Yang, K., and Zondlo, M.: A physics-based approach to oversample multi-satellite,

multi-species observations to a common grid, Atmospheric Measurement Techniques Discussions, 1-30, doi:10.5194/amt-2018-253, 2018.

Valin, L. C., Russell, A. R., Hudman, R. C., and Cohen, R. C.: Effects of model resolution on the interpretation of satellite NO₂ observations, Atmospheric Chemistry and Physics, 11, 11647-11655, doi:10.5194/acp-11-11647-2011, 2011.

30   Valin, L. C., Russell, A. R., and Cohen, R. C.: Variations of OH radical in an urban plume inferred from NO₂ column measurements, Geophysical Research Letters, 40, 1856-1860, doi:10.1002/grl.50267, 2013.

Walcek, C. J.: Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection, Journal of Geophysical Research: Atmospheres, 105, 9335-9348, doi:10.1029/1999jd901142, 2000.

Yan, Y.-Y., Lin, J.-T., Chen, J., and Hu, L.: Improved simulation of tropospheric ozone by a global-multi-regional two-way coupling model system, Atmospheric Chemistry and Physics, 16, 2381-2400, doi:10.5194/acp-16-2381-2016, 2016.

Zhang, R., Lin, J., and Zhang, Q.: Satellite-based evaluation of nitrogen oxides emission inventory for Beijing at a high resolution, Bachelor, Peking University, unpublished, 19 pp., 2014.

5 Zhao, Y., Nielsen, C. P., Lei, Y., McElroy, M. B., and Hao, J.: Quantifying the uncertainties of a bottom-up emission inventory of anthropogenic atmospheric pollutants in China, Atmospheric Chemistry and Physics, 11, 2295-2308, doi:10.5194/acp-11-2295-2011, 2011.

Zheng, B., Zhang, Q., Tong, D., Chen, C., Hong, C., Li, M., Geng, G., Lei, Y., Huo, H., and He, K.: Resolution dependence of uncertainties in gridded emission inventories: a case study in Hebei, China, Atmospheric Chemistry and Physics, 17, 921-933, doi:10.5194/acp-17-921-2017, 2017.
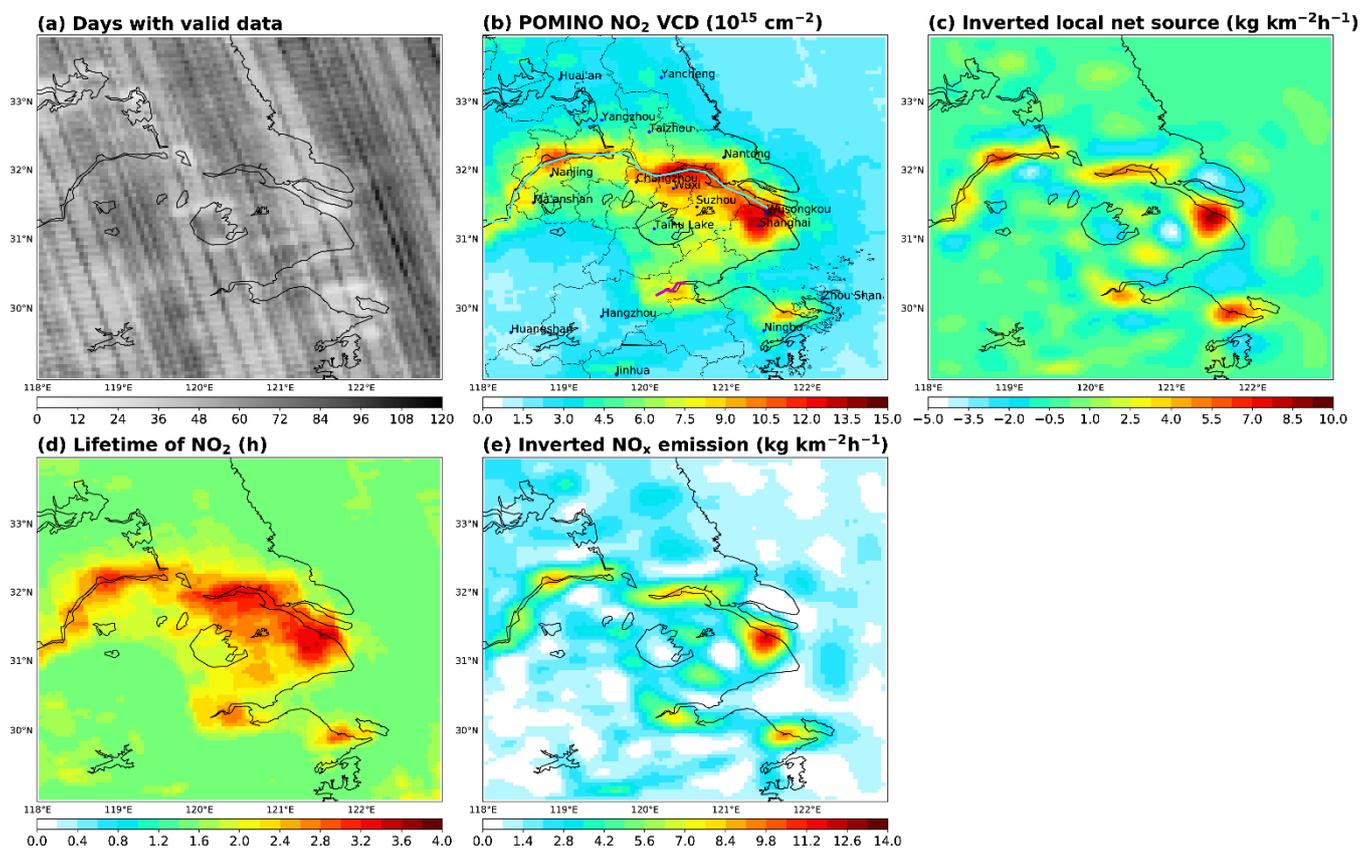
Figure 1. (a) Number of days with valid data in each grid cell over summer months of 2012–2016. (b) POMINO NO_2 VCDs averaged over summer 2012–2016. Cities and locations mentioned in this paper are denoted. The Yangtze River is marked as a blue line, and the Qiantang River (passing Hangzhou) a purple line. (c) Inverted local net source. (d) Inverted NO_x lifetime due to both chemical loss and deposition. (e) Inverted NO_x emissions from anthropogenic and soil sources.
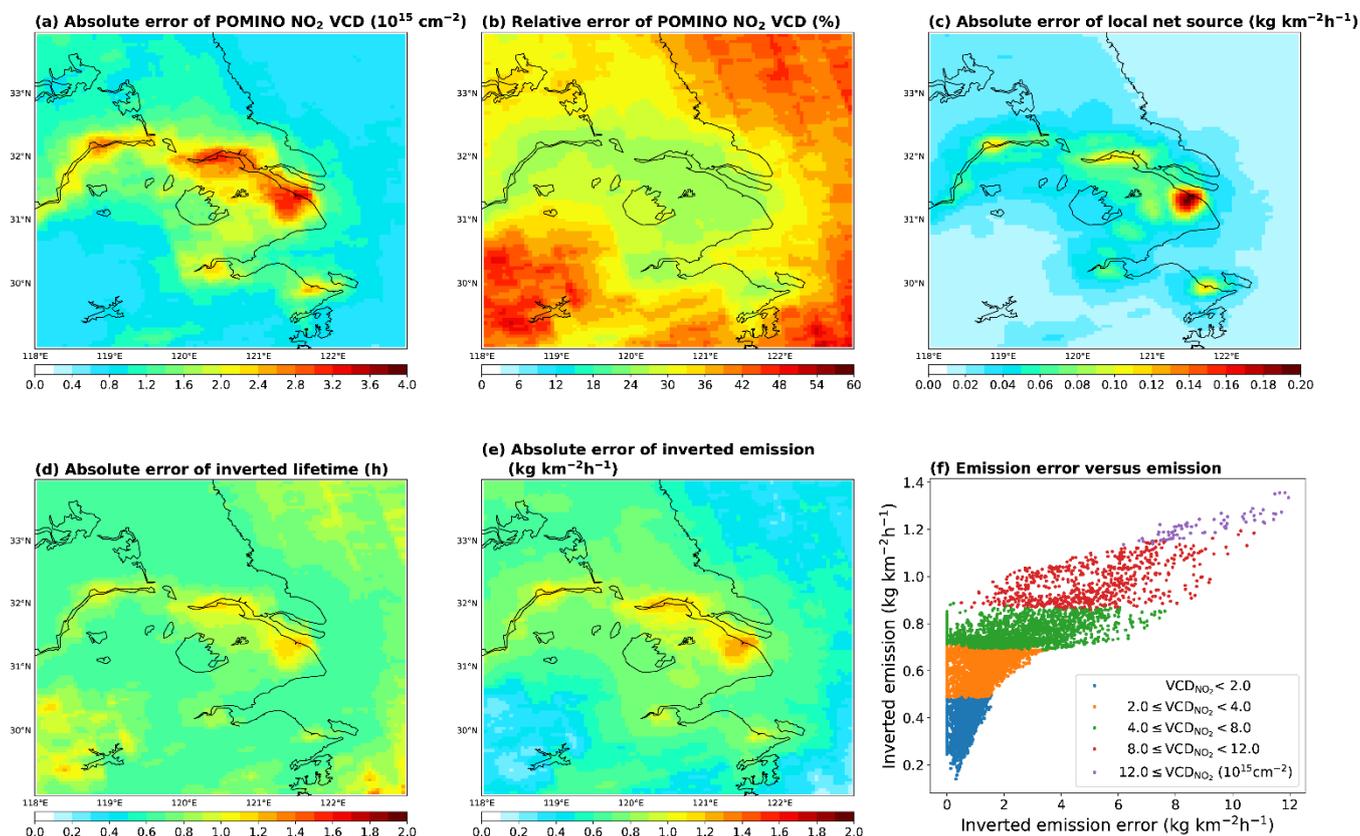
Figure 2. (a) Absolute error (1σ) of POMINO NO$_2$ VCD at each grid cell at $0.05° \times 0.05°$ resolution. (b) Relative error (1σ) of POMINO NO$_2$ VCD. (c) Absolute error (1σ) of the inverted local net source. (d) Absolute error (1σ) of the inverted NO$_x$ lifetime due to chemical loss and deposition. (e) Absolute error (1σ) of inverted NO$_x$ emission. (f) Absolute error (1σ) of inverted NO$_x$ emission as a function of inverted NO$_x$ emission at each grid cell. Data points are coloured according to the magnitudes of POMINO NO$_2$ VCDs.
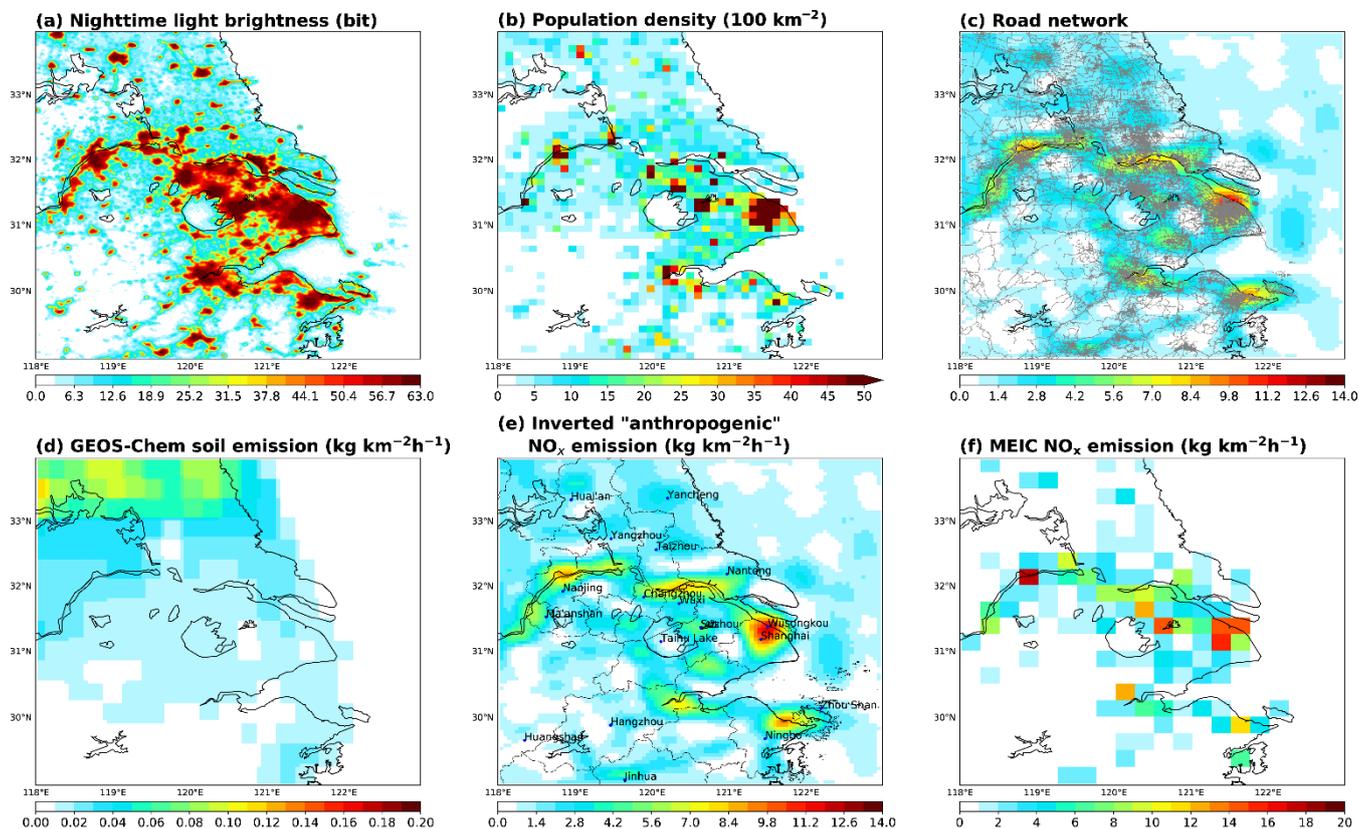
Figure 3. (a) Nighttime light brightness on a $0.5' \times 0.5'$ grid in 2012. (b) Population density averaged over 2012–2016 on a $0.1° \times 0.1°$ grid. (c) Road network (grey lines) superimposed upon the inverted NO_x emissions. (d) Soil NO_x emissions calculated by a nested GEOS-Chem simulation ($0.25° \times 0.3125°$) and regraded to the $0.05° \times 0.05°$ grid. (e) Inverted "anthropogenic" NO_x emissions on a $0.05° \times 0.05°$ grid. City names and boundaries are given. (f) MEIC NO_x emissions on a $0.25° \times 0.25°$ grid.
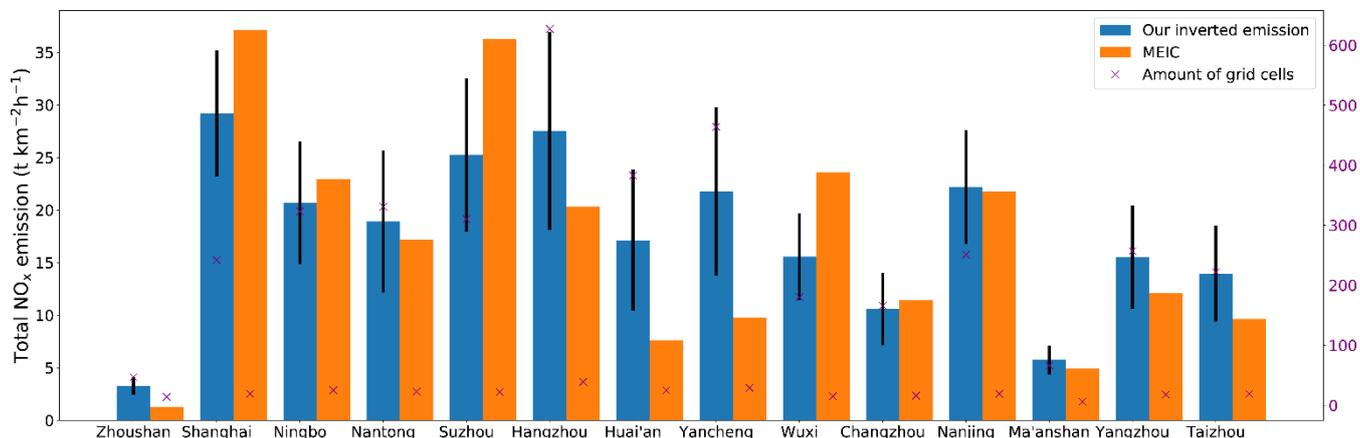
Figure 4. Total anthropogenic NO_x emission in each city for summer 2012–2016 inverted here versus from the MEIC inventory. Soil NO_x emissions calculated by nested GEOS-Chem have been subtracted from our inverted emission data. Black vertical lines denote the uncertainty (1σ). Red points denote the numbers of grid cells covered by each city at different resolutions (0.05° × 0.05° in our inverted dataset and 0.25° × 0.25° in MEIC).
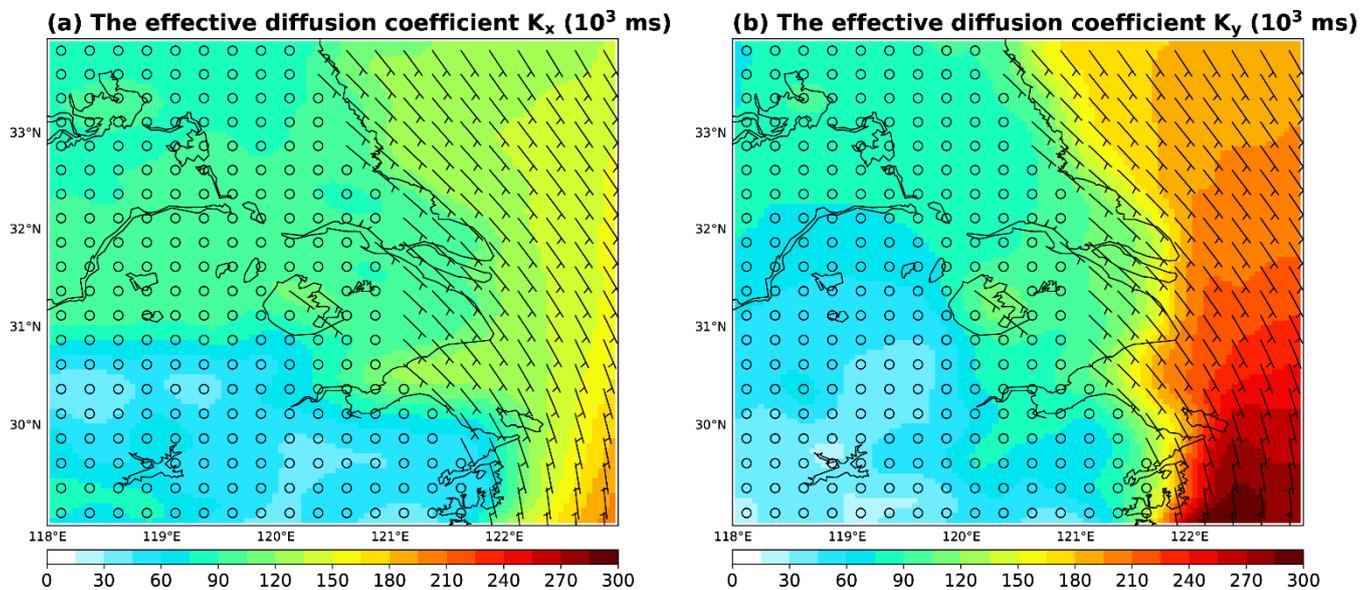
Figure A1. The effective diffusion coefficients for summer 2012–2016 on a 0.05° × 0.05° grid. Overlaid is the temporal mean wind vector, which is plotted for every $5 \times 5 = 25$ grid cells to enhance the readability.
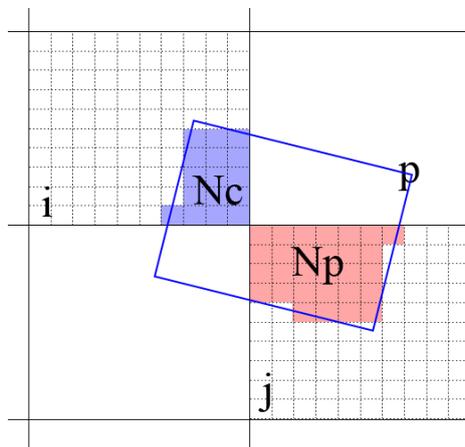
Figure B1. Illustration of how the original PHLET model grid cell i is projected to the satellite pixel p and then to the final grid cell j through the SCM approach. The size of the satellite pixel is scaled down to be comparable with the size of a model grid cell, for illustration purpose.
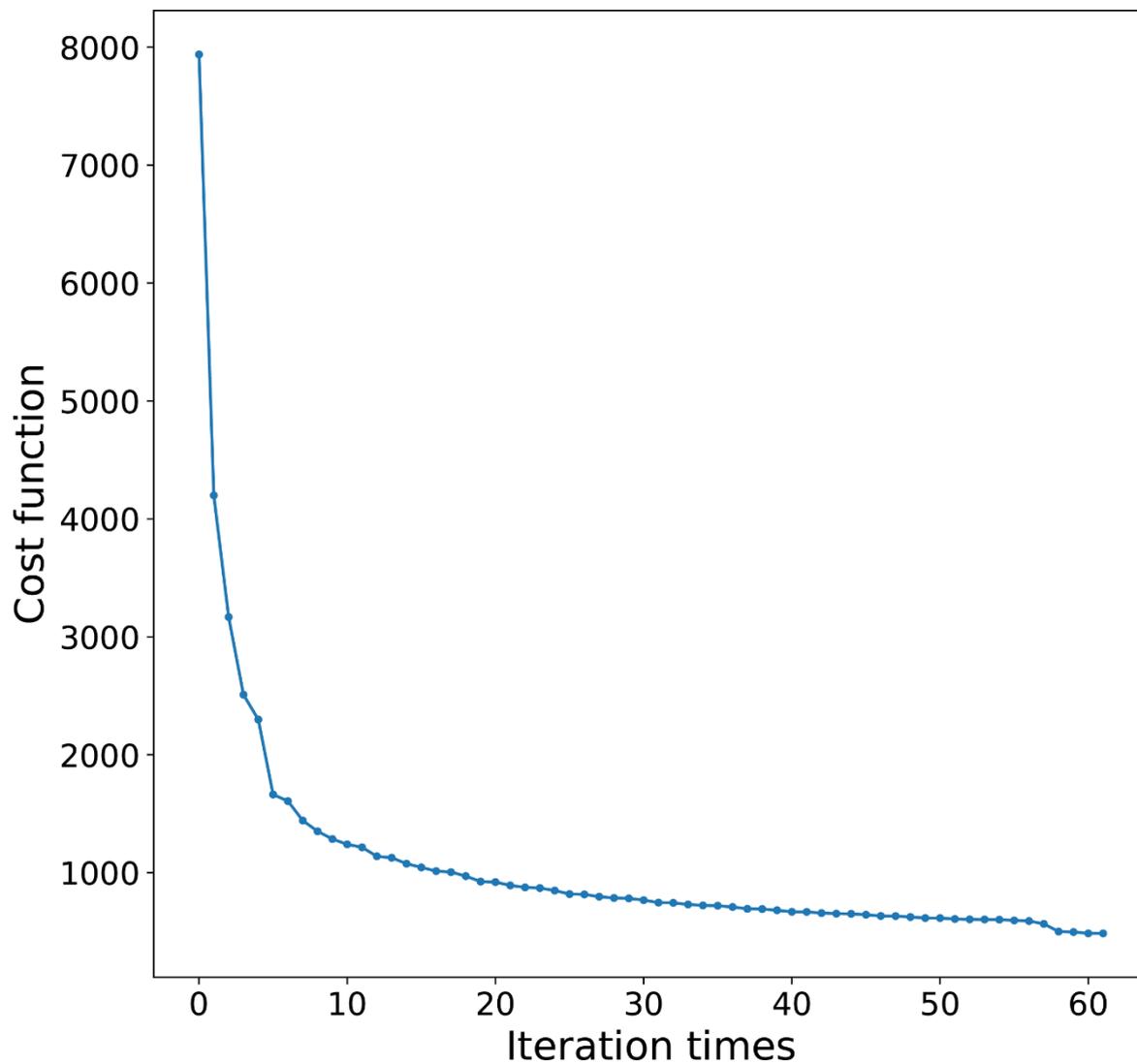
Figure C1. Variation of the cost function *J* with iteration times.
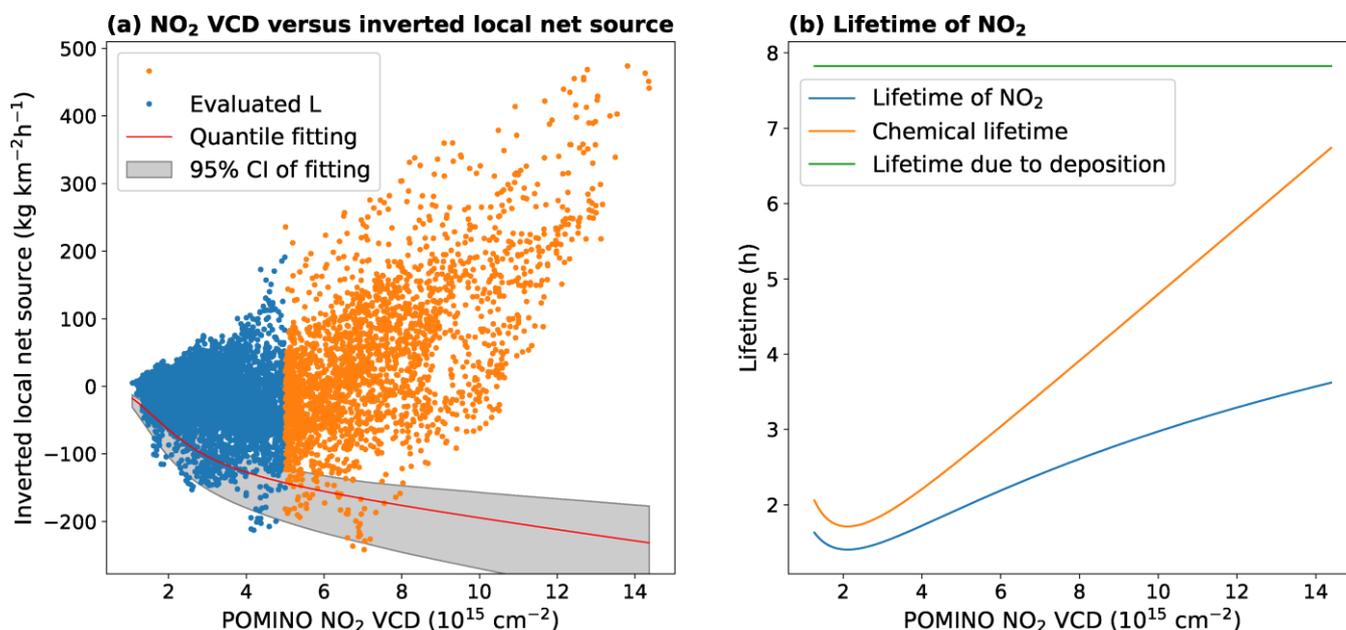
Figure D1. (a) Scatter plot for POMINO NO_2 VCDs (x-axis) and inverted local net sources (y-axis) across individual grid cells. Grid cells with NO_2 VCDs below (above) 5×10^{15} molecules cm^{-2} are coloured in blue (orange). The red line and shading denote the median and uncertainty (95% CI) of the quantile fitting, respectively, to estimate the nonlinear relationship between NO_2 VCD and lifetime, based on data in the low-emission areas. (b) The derived relationship between NO_2 VCD and lifetime across the range of NO_2 VCDs in the YRD area.
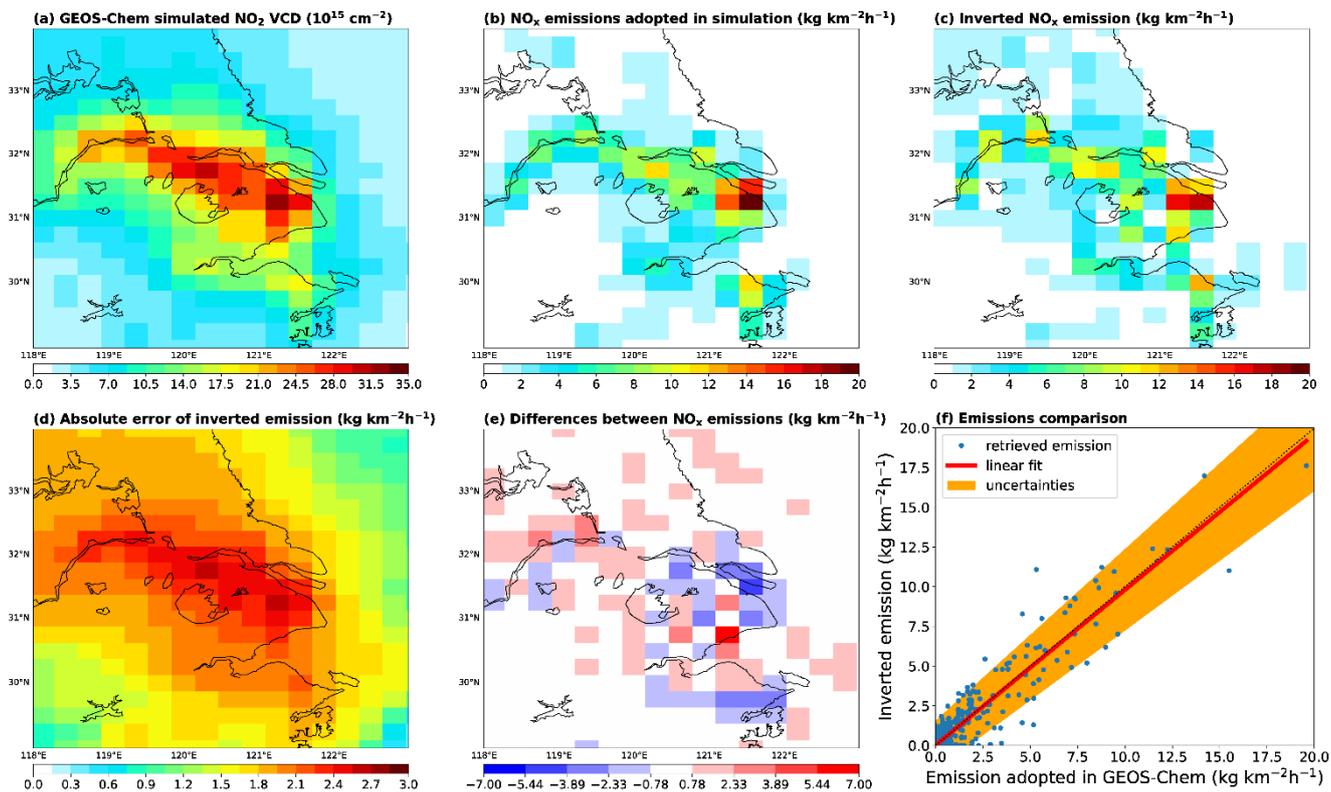
Figure E1. (a) GEOS-Chem simulated NO_2 VCDs at the 0.3125° × 0.25° resolution for summer 2014. (b) Anthropogenic NO_x emissions used in GEOS-Chem. (c) Inverted anthropogenic emissions based on GEOS-Chem simulated NO_2 VCDs. (d) Absolute errors (1σ) of inverted emission data. (e) Differences between the inverted emissions and GEOS-Chem emissions (inverted minus GEOS-Chem). (f) Scatter plot for the inverted emissions (y-axis) and GEOS-Chem emissions (x-axis). The red line represents least square linear fitting. The shading represents the fitting by accounting for errors in the inverted emission data, i.e., inverted emissions + 1σ for the upper bound, and inverted emissions – 1σ for the lower bound. The black dotted line denotes the 1:1 line.