



Constraining emission estimates of carbon monoxide using a perturbed emissions ensemble with observations: a focus on Beijing

Le Yuan¹ · Olalekan A.M. Popoola¹ · Christina Hood² · David Carruthers² · Roderic L. Jones¹ · Huan Liu³ · Zhaofeng Lv³ · Qiang Zhang⁴ · Alexander T. Archibald^{1,5}

Received: 28 January 2021 / Accepted: 2 May 2021 / Published online: 15 May 2021
© The Author(s) 2021

Abstract

The reliability of air quality simulations has a strong dependence on the input emissions inventories, which are associated with various sources of uncertainties, particularly in regions undergoing rapid emission changes where inventories can be ‘out of date’ almost as soon as they are compiled. This work provides a new methodology for updating emissions inventories by source sector using air quality ensemble simulations and observations from a dense monitoring network. It is adopted to determine the short-term trends in carbon monoxide (CO) emissions, an important pollutant and precursor to tropospheric ozone, in a study area centred around Beijing following the implementation of clean air policies. We sample the uncertainties associated with using an a priori emissions inventory for the year 2013 in air quality simulations of 2016, using an atmospheric dispersion model combined with a perturbed emissions ensemble (PEE), which is constructed based on expert-elicited uncertainty ranges for individual source sectors in the inventory. By comparing the simulation outputs with observational constraints, we are able to constrain the emissions of key source sectors relative to those in the a priori emissions inventory. From 2013 to 2016, we find a 44–88% reduction in the transport sector emissions ($0.92\text{--}4.4 \times 10^5$ Mg in 2016) and a minimum 61% decrease in residential sector emissions ($<3.5 \times 10^5$ Mg in 2016) within the study area. We also provide evidence that the night-time fraction of traffic sources in 2016 was higher than that in the 2013 emissions inventory. This study shows the applicability of PEEs and high-resolution observations in providing timely updates of emission estimates by source sector.

Keywords Air quality model · Emissions inventory · Emissions uncertainty · Ensemble simulations · Low-cost sensors · Network measurements

Introduction

Emissions inventories are essential inputs to air quality modelling from global to local scales (Granier et al. 2011; Hoesly et al. 2018). Pollutant concentrations modelled with these emissions are often used to assess the disease and mortality burden (Lelieveld et al. 2015; Silva et al. 2017) or the loss in crop yields (Tai et al. 2014; Solazzo et al. 2018) caused by pollution exposure, and serve as the basis for pollution control legislation and practice. However, emissions inventories are subject to various types of uncertainties and many studies have demonstrated that such uncertainties can be propagated through the modelled concentrations into biases in the health or vegetation damage estimates. For example, Crippa et al. (2019) estimated 2.1 million premature deaths related to PM_{2.5} pollution globally in 2010, with an uncertainty range of ± 1.1 million solely due to emission uncertainty, which is comparable to the uncertainties associated with air quality

✉ Le Yuan
ly299@cam.ac.uk

¹ Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK

² Cambridge Environmental Research Consultants, Cambridge, UK

³ State Key Joint Laboratory of ESPC, State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, International Joint Laboratory on Low Carbon Clean Energy Innovation, School of the Environment, Tsinghua University, Beijing, China

⁴ Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China

⁵ National Centre for Atmospheric Science, Cambridge, UK

models (Liang et al. 2018) and exposure-response functions (Cohen et al. 2017). Hence, it is crucial to quantify and constrain the uncertainties in emissions inventories.

Emissions inventories can be developed following two broad approaches. The bottom-up approach calculates emissions from each source sector and fuel type by combining specific emission factors and corresponding activity rates (Liu et al. 2015). It provides high spatiotemporal detail but requires a large amount of input data which are likely to contain errors of various kinds. For instance, several studies quantified the uncertainties in emission factors (Zhao et al. 2012) and energy statistics (Hong et al. 2017) for China, the world's largest energy consumer and one of the top emitters of air pollutants, the accuracy of whose emission estimates is of global importance. The top-down proxy-based approach downscales and allocates national or regional level statistics into a grid using spatial proxies representing local activity levels (Raupach et al. 2010). It is less data-intensive but may introduce additional uncertainties into the distributions of emissions. For example, Geng et al. (2017) demonstrated that with the same total emissions, modelled NO₂ columns for China are sensitive to the spatial surrogates used. Moreover, the correlation between emissions and a proxy is often sensitive to the spatial resolution and tends to break down at finer scales (Oda and Maksyutov 2011). A single inventory can be established incorporating both approaches (e.g. Zheng et al. 2017), as emission factors and activity data may not be available for all source sectors and the use of proxies is essential for gridding; thus, different and complex types of uncertainties may co-exist.

In addition to the inherent uncertainties in emission estimates, there are uncertainties introduced by the trends of emissions. This arises from the fact that emissions inventories are inevitably time lagged — they are inventories of past emissions. Depending on the approach(es) adopted, the lag between the occurrence of emissions and the completion of an inventory may vary between one and several years (Janssens-Maenhout et al. 2015). Yet inventories are not only applied in hindcasting pollutant concentrations for the period when the emissions occurred (Colette et al. 2017), but are also widely used in forecasting (Brasseur et al. 2019). This can be particularly problematic in regions undergoing rapid emission changes such as the case of China. During the past decade, China has been actively combating its air pollution problem through a series of clean air policies (Cheng et al. 2020). The most up-to-date assessment identified nationwide reductions in SO₂, NO_x, CO, PM₁₀ and PM_{2.5} emissions between 2010 and 2017, while NMVOC and NH₃ emissions increased and plateaued, respectively (Zheng et al. 2018). Many provincial and city-level governments have also been implementing their own pollution control measures which are of differing stringency and effectiveness, leading to spatially inhomogeneous emission changes (Cheng et al. 2019). Hence, when an existing emissions inventory is used to simulate air quality at

a later point in time, the uncertainties and errors due to emission trends should also be taken into consideration.

Given these uncertainties, there have been numerous studies that verify and/or optimise emissions inventories. The eddy covariance technique, originally established to quantify the biosphere-atmosphere carbon and heat exchange (Baldocchi 2003), is also increasingly used to measure anthropogenic pollutant fluxes from urban canopies. Vertical fluxes provide a direct measure of the emission strength from an upwind source area referred to as the flux footprint, which can be scaled for comparison with emissions inventories (Lee et al. 2015; Squires et al. 2020). To expand the small footprint areas associated with tower-based measurements, measurements from mobile platforms such as towers on vehicles (Moore et al. 2009) and aircrafts (Vaughan et al. 2016) have been used to obtain urban-scale estimates. Another key approach is to infer emission strengths from ambient pollutant concentrations (Tang et al. 2013; Carruthers et al. 2019) but more commonly satellite retrievals of total column quantities due to their wider spatiotemporal coverage (Wang et al. 2010, 2015; Liu et al. 2016). Statistical inversion techniques span complex fitting algorithms and mass balance box modelling (de Foy et al. 2014). Other studies have optimised simulations of chemical transport models using variational or sequential algorithms (Streets et al. 2013).

In this study, uncertainties in an emissions inventory are examined using forward modelling with a perturbed emissions ensemble (PEE), analogous to the perturbed physics ensemble (or perturbed parameter ensemble) in climate modelling. Like air quality simulations, climate simulations are subject to various sources of uncertainty, including uncertainties in the initial conditions and formulations of models and natural climate variability (Palmer 2000). Ensemble approaches have been widely adopted to quantify such uncertainties. The Intergovernmental Panel on Climate Change Assessment Reports use multi-model ensembles, i.e. a collection of simulations that sample a wide range of model structural choices and internal variability performed by different modelling centres. Conversely, the perturbed physics ensemble can be used to systematically sample the parameterisation uncertainty within a single model (Collins et al. 2011). Perturbations can be made to one parameter at a time (Murphy et al. 2004). However, any effects of parameter interactions are then ignored. Moreover, the fraction of the parameter space which can be explored rapidly reduces with growing number of parameters (Saltelli and Annoni 2010). To allow for parameter interaction and to increase the sampling efficiency, many studies have simultaneously perturbed multiple parameters (Lambert et al. 2013). The uncertainty range for each parameter is commonly determined in consultation with experts in the field (Brierley et al. 2010). Lee et al. (2013) described rigorous procedures of the Sheffield Elicitation Framework (Oakley and O'Hagen 2019).

This study aims to reduce the uncertainties in carbon monoxide (CO) emissions in an a priori emissions inventory for Beijing in the light of recent developments. Apart from being a major pollutant, CO plays an important role in atmospheric chemistry as the dominant sink for hydroxyl radicals and a precursor of tropospheric ozone. CO concentrations in Beijing have showed a continuous downward trend since 2013 (the year for which the inventory was compiled) (Li et al. 2020), which has been attributed to reductions in emissions locally (Xue et al. 2019) and in the wider region (Jiang et al. 2017). This could result in large uncertainties when the emissions inventory is used in simulations of more recent years. By perturbing key parameters in the inventory within expert-elicited uncertainty ranges, a PEE was constructed. It was subsequently used in an atmospheric dispersion model to simulate CO concentrations. By comparing the simulation outputs with observations, the initial uncertainty ranges could be constrained and the trends in emissions could be detected.

Data and methods

Base emissions

A special version of the Multi-resolution Emission Inventory for China (MEIC) v1.3, developed by Tsinghua University (<http://www.meicmodel.org>, last accessed 22 August 2020), was developed for the UK-China joint research programme Atmospheric Pollution and Human Health in a Chinese Megacity (APHH-Beijing) (Shi et al. 2019) and used in this work. Details of the data and approaches used in developing the MEIC can be found in Zheng et al. (2018). This version covers an area of 120 km × 150 km which extends over Beijing (excluding most of Yanqing, Huairou, Miyun Districts and the westernmost parts of Mentougou and Fangshan Districts) and regions in Hebei Province to the east and south with 3 km × 3 km horizontal resolution. For this study, a smaller region of 35 × 48 grid cells was cropped (starting from the northwest corner) and used as the a priori emissions to be optimised, hereinafter referred to as the base emissions (Fig. 1). There are seven vertical layers with the tops of layers at 38, 90, 152, 228, 337, 480 and 660 m above ground, respectively. This version provides emissions of CO, NO, NO₂, VOC, SO₂, PM₁₀ and PM_{2.5} from four sectors (industry, power, residential and transport) in the year 2013. Each sector is characterised by a specific hourly diurnal profile applied to all pollutants. Emissions are monthly varying with no day-of-week variation. The sectors also differ in their vertical distributions, with residential and transport sectors emitting solely in the lowest layer. Industrial sources are distributed within the lowest three layers. Power emissions are present in all except the lowest layer.

Observations

Observed concentrations were used to evaluate the PEE simulations to constrain uncertainty ranges of the base emissions. The Beijing Municipal Environmental Monitoring Center (BJMEMC) currently operates 35 long-term air quality monitoring sites, 33 of which were in operation in 2016 and located within the study area (Fig. 1, Table S1), determined by the extent of the base emissions. The sites can be categorised into five types according to their local environment: urban sites in the 6 central districts, suburban sites in the outer district centres, traffic monitoring sites near major roads, regional background sites that sample pollution from surrounding regions and one clean site (Chen et al. 2015). All sites were equipped with continuous automated monitoring systems for CO, NO₂, O₃, SO₂, PM₁₀ and PM_{2.5}. The CO analysers are required to have an indication error within ±2% of the full measurement range (0–58 mg/m³) (Ministry of Environmental Protection of the People's Republic of China 2013). Real-time hourly concentrations are published on the website of BJMEMC, while historical data is not openly available. The provisional real-time data is archived by several providers. We used 2013 data from <http://data.epmap.org> (last accessed on 1 April 2021) and 2016 data available at <https://quotsoft.net/air/> (last accessed on 22 August 2020).

A second dataset was established using low-cost SNAQ (Sensor Network for Air Quality) boxes during the APHH-Beijing winter (November–December 2016) and summer campaigns (May–June 2017) (Shi et al. 2019). The SNAQ boxes were equipped with electrochemical sensors and optical particle counters measuring CO, CO₂, NO, NO₂ and particulate matter every 20 s (Popoola et al. 2018). A total of 21 sensor nodes were deployed around Beijing for near-surface measurements in a variety of microenvironments with one node outside of the modelling domain (Fig. 1, Table S2). This dataset has been validated against reference measurements during the campaigns and data from the long-term monitoring sites (Shi et al. 2019). The CO measurements are accurate to within ±10% for concentrations in the range of 0–10 mg/m³.

Model

ADMS-Urban, a state-of-the-art high-resolution, urban-scale, quasi-Gaussian dispersion model (McHugh et al. 1997; Owen et al. 2000), was used for the PEE simulations. It has been used to produce fine-scale air quality simulations in many cities worldwide (e.g. Hood et al. 2018), which can be further applied in assessing pollution exposure and health risk (He et al. 2019), or testing the effectiveness of emission control strategies (Taksibi et al. 2020). In Beijing, it has also been adopted to model air quality at the street level (Biggart et al. 2020). Gaussian dispersion models assume a Gaussian concentration

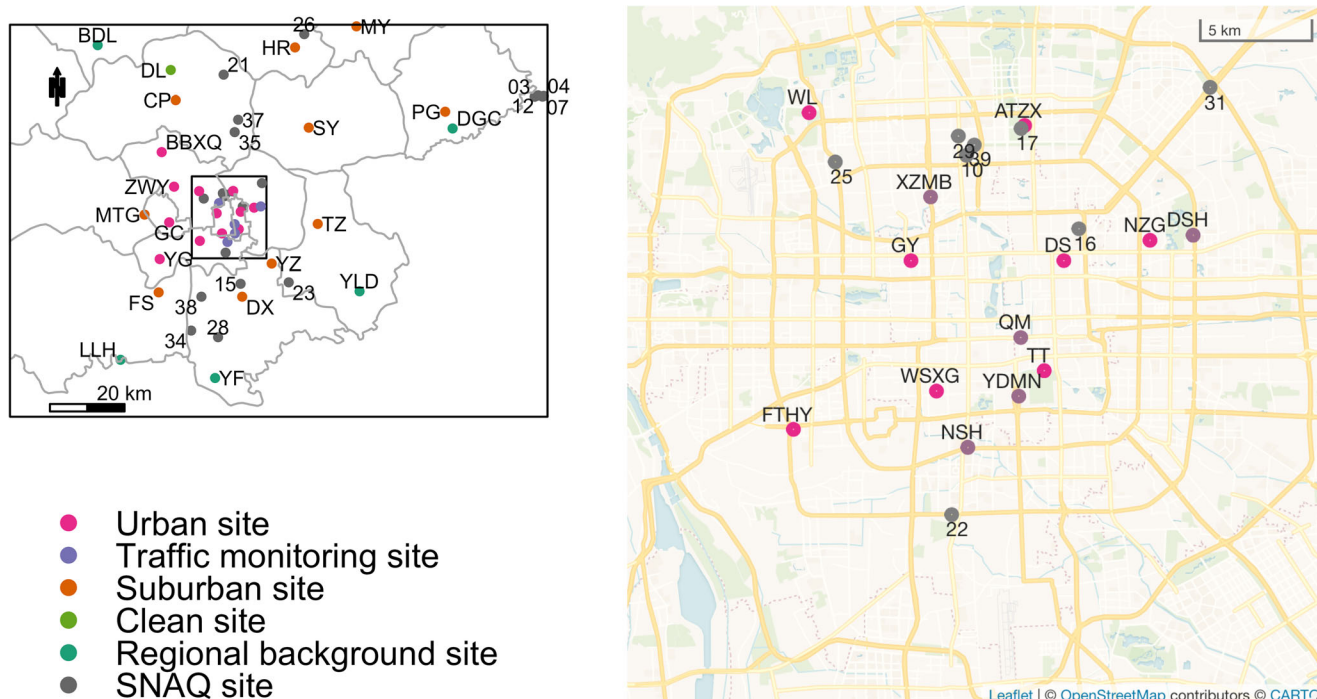


Fig. 1 The modelling domain as set by the extent of the base emissions with the locations of observations, including a zoomed view of those within the 5th Ring Road in Beijing. Long-term monitoring sites are colour-coded according to the site type and labelled by their acronyms.

Locations of SNAQ measurements are coloured in grey. Coordinates and full names of the long-term sites are listed in Table S1. Coordinates of the SNAQ sites can be found in Table S2. The administrative divisions of Beijing are shown by light grey contours

distribution in the vertical and crosswind directions in a plume downwind of a source (Arya 1999). The plume spread parameters (i.e. Gaussian distribution parameters) depend on the conditions of the atmospheric boundary layer which are described in ADMS-Urban using the Monin-Obukhov similarity theory (Venkatram 1996). The minimum meteorological input data required are wind direction and speed and at least one of the Obukhov length (L_O), the surface sensible heat flux or the cloud cover, based on which the atmospheric boundary layer height (ABLH) is estimated. The model uses the ratio of ABLH to L_O to generate continuous profiles of boundary layer properties such as the mean wind speed, the turbulent velocity and the buoyancy frequency, which are then used to calculate plume spread parameters that also vary with the source and plume height. Under convective conditions (here defined as $ABLH/L_O < -0.3$), a positively skewed concentration distribution is assumed in the vertical (Cambridge Environmental Research Consults Limited 2017). ADMS-Urban is integrated with fast chemistry schemes that simulate the NO_x - O_3 -VOC chemistry and the sulphate chemistry (Azzi et al. 1992; Malkin et al. 2016). Chemical production and loss of CO are not simulated. Due to its long lifetime of 1–3 months (Khalil and Rasmussen 1990) and the small contributions from the photochemical production via VOC oxidation relative to the primary emissions, CO is commonly modelled as an inert species subject to dispersion and mixing only at city to regional scales (Saide et al. 2011; Panagi et al. 2020).

Expert elicitation of emissions uncertainties

To explore the effect of varying magnitudes of source sectors on CO concentrations, individual emission parameters were perturbed simultaneously. The choice of parameters was based on the structure of the base emissions (Table 1). Three-dimensional emissions were split into ground (the lowest layer) and elevated emissions (all upper layers) to allow for perturbations to the vertical emissions profile. As power sector emissions are present in all but the lowest layer, they were represented by a single parameter and their vertical distribution was unaffected. In addition to perturbing the magnitude of total transport sector emissions, another parameter was introduced to represent the night-time fraction of traffic sources, defined as those occurring during 11pm–6 am (inclusive) following Biggart et al. (2020).

Uncertainty ranges of the selected parameters were determined through expert elicitation. Participants included researchers who also used the same emissions inventory and experts specialised in developing emissions inventories for the region. They were invited to anonymously suggest a lower and an upper bound of uncertainty for each parameter, such that the true parameter value would be unlikely to fall outside of this range. The range represents an uncertainty range of the emission trend, i.e. the ratio of 2016 emissions to the base emissions from 2013, for all parameters except the night-time fraction of transport sources, for which an uncertainty

Table 1 Selected emission parameters^a and the respective uncertainty ranges sampled by the initial and the optimised perturbed emissions ensembles (PEEs)

Parameter	Initial PEE		Optimised PEE	
	Min	Max	Min	Max
Industry sector ground level CO emissions	0.3	1.8	0.1	1.8
Industry sector elevated CO emissions	0.3	1.4	0.1	1.4
Power sector CO emissions ^b	0.2	1.5	NA	NA
Power sector CO emissions below 152 m	NA	NA	0.1	1.8
Power sector CO emissions above 152 m	NA	NA	0.1	1.8
Residential sector CO emissions	0.3	1.5	0.1	1.5
Transport sector CO emissions	0.4	2	0.1	1.5
Night-time fraction of transport sector CO emissions	0.1	0.5	0.1	0.3

^a The night-time fraction of transport sector CO emissions is defined as a proportion (%) of the daily totals. Other parameters are defined as ratios of the 2016 emissions to the base emissions from 2013

^b Power sector CO emissions in the initial PEE are split into two parameters—power sector CO emissions below and above 152 m in the optimised PEE

range of the absolute value was sought. The responses from the first round were sent back anonymously to the experts for review. The maximum and minimum values given by all participants for a parameter in the second round of responses were used to determine its uncertainty range (Table 1, column *Initial PEE*). Inherent uncertainties in the base emissions were not considered as they are mostly likely small compared to the uncertainties due to emission reductions. Zhao et al. (2012) estimated an uncertainty range of -20% to $+45\%$ for CO emissions in China in 2005, which was dominated by uncertainties in emission factors. Another study quantified a 15.6% uncertainty associated with energy statistics for CO₂ (co-emitted with CO) emissions in 2012 (Hong et al. 2017). Though not calculated for the base emissions in this study, these values are substantially smaller than the uncertainty ranges in Table 1.

Following the same procedure, six additional parameters and their uncertainty ranges were determined for NO_x emissions which will be constrained in a separate paper. Together the parameters constitute a 12-dimensional parameter space. As CO is unreactive in ADMS-Urban, varying NO_x emissions have no impact on the modelled CO concentrations (and vice versa). Hence, all parameters can be perturbed simultaneously. To efficiently probe this uncertainty space with a limited number of simulations, the maximin Latin hypercube sampling was used. A Latin hypercube design divides the range of each parameter into n intervals of equal probability and samples once from each interval (McKay et al. 1979). There can be a large number of permutations, but some are not optimal, for instance, when parameters are highly correlated or when the design leaves large areas of the space unexplored. A secondary design criterion is thus needed (Joseph and Hung 2008). The maximin distance criterion is a space-filling design that maximises the minimum distance between pairs of sample points (Johnson et al. 1990). Following the $n =$

$10d$ (d is the number of parameters) rule of thumb (Loeppky et al. 2009) adopted in statistical emulation studies, 120 samples were drawn. Using the sampled parameter values, a spatially uniform scaling was applied to the base emissions to produce an initial 120-member PEE.

Simulation setup and initial evaluation

The initial 120-member PEE and the base emissions were used to simulate hourly CO concentrations at all observation locations (see Fig. 1) in 2016. All simulations were forced with the same lateral boundary conditions including meteorology and background pollution concentrations.

Hourly average wind direction and speed, cloud cover and air temperature observed at the Beijing Capital International Airport (located in the northeast of the modelling domain) in 2016 were obtained from the NOAA Integrated Surface Database (Smith et al. 2011). While these near-surface conditions were first assumed to apply throughout the modelling domain, local perturbations due to friction against obstacles such as buildings were accounted for through the additional definition of the roughness length (z_0) and the minimum L_0 . Based on previous literature (e.g. Stewart and Oke 2012), a z_0 of 0.5 m and a minimum L_0 of 30 m were used for the meteorological measurement site in open landscape. For the output locations mostly in more built-up, yet still heterogeneous environments as revealed by the site types, the z_0 and minimum L_0 were set to 1 m and 100 m, respectively. The z_0 of 1 m results in a minimum ABLH of 120 m imposed by the model (Cambridge Environmental Research Consults Limited 2017), which is broadly consistent with ceilometer measurements during the APHH-Beijing campaigns (Kotthaus and Grimmond 2018; Shi et al. 2019).

Background pollutant concentrations are required as an input for ADMS-Urban to represent pollution levels not

explicitly modelled (i.e. not included in the input emissions) and, for reactive species, to constrain the rate coefficients of reactions (Cambridge Environmental Research Consults Limited 2017). A baseline CO concentration from the 33 long-term monitoring sites within the modelling domain was used as the background. It was defined as the 10th percentile of all measurements in a moving 3-h window to reflect the expected smooth variation of a baseline (Ruckstuhl et al. 2012) and to account for any measurement errors (Popoola et al. 2018).

An evaluation of the initial PEE simulations' performance in modelling local CO concentrations observed at the long-term monitoring sites in 2016 is provided in Fig. 2. Local concentrations were calculated by subtracting the input background levels from both the modelled and the observed total signals. Based on the definition of background concentrations and the fact that CO does not undergo chemical reactions in ADMS-Urban, it is evident that local CO concentrations represented concentrations resulting solely from the input CO emissions, which were to be constrained.

Figure 2 shows the performance of the initial PEE simulations and two simulations using the base emissions for the year 2013 (hereinafter referred to as the 2013 run) and the year 2016 (hereinafter referred to as the base run). The 2013 run was forced with meteorology and background concentrations from 2013, the year for which the base emissions were compiled. The modelled CO concentrations showed good agree-

ment with those observed in 2013 overall, with larger discrepancies at a few sites during the autumn and winter months (Fig. S1). This validated the performance of ADMS-Urban given correct inputs, for example, an up-to-date emissions inventory. The base run was forced with the base emissions, the same meteorology and background concentrations from 2016 as those input in the initial PEE simulations and was also evaluated against observations from 2016. For all simulations, we calculated the normalised mean bias factors (NMBF) in local CO concentrations defined as:

$$\begin{aligned} NMBF &= \frac{\overline{Mod}}{\overline{Obs}} - 1, \text{ if } \overline{Mod} \geq \overline{Obs} \\ &= 1 - \frac{\overline{Obs}}{\overline{Mod}}, \text{ if } \overline{Mod} < \overline{Obs} \end{aligned} \quad (1)$$

where \overline{Mod} and \overline{Obs} are the modelled and observed annual mean local CO concentrations, respectively. A positive NMBF indicates an overestimation by a factor of NMBF+1, while a negative NMBF represents an underestimation by a factor of 1–NMBF (Yu et al. 2006). The NMBFs showed that the degree of overestimation at most urban and traffic monitoring sites in the base run was substantially reduced in the 2013 run, indicating that the base emissions compiled for the year 2013 became high biased in 2016, which was expected given the series of clean air policies implemented.

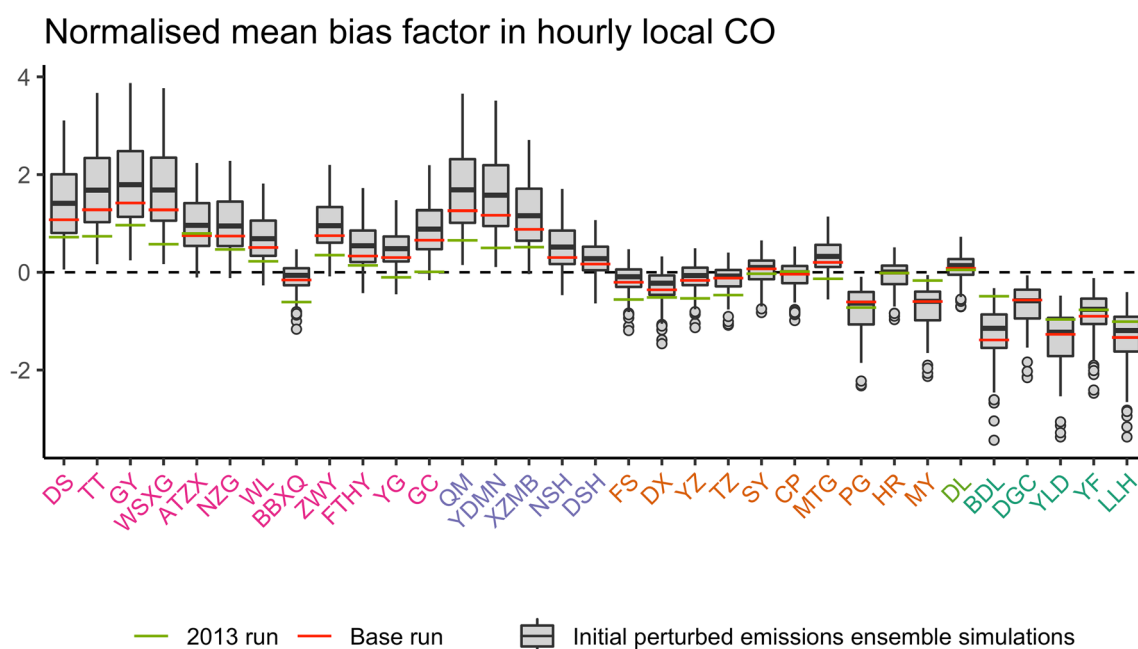


Fig. 2 Distribution of the normalised mean bias factors (NMBFs) in hourly local (i.e. background removed) CO concentrations calculated for the initial perturbed emissions ensemble (PEE) simulations (shown as light grey box-and-whisker plots), the base run (red line segments) and the 2013 run (green line segments) at each long-term monitoring site. The

monitoring sites are colour-coded according to the site type: urban site (magenta), traffic monitoring site (purple), suburban site (orange), clean site (light green) and regional background site (green). Performance of the 2013 run could not be evaluated at the sites NSH, DSH and DGC, where CO observations from 2013 were unavailable

The initial PEE simulations also resulted in a widespread overprediction of local CO concentrations. All simulations overestimated the annual mean local CO concentrations at 6 urban and traffic monitoring sites (Fig. 2). In other words, the ensemble failed to simulate concentrations comparable to the observations at these locations; thus, the uncertainty ranges of emissions could not be constrained. In addition to the overestimated mean concentrations, the modelled mean diurnal variations of local CO concentrations were characterised by two distinct peaks in the mornings and evenings across the modelling domain. The observed diurnal profiles were much flatter at most sites, with other locations showing elevated concentrations throughout the night (Fig. S2). This suggested that the diurnal distribution of CO emissions, most likely that from traffic sources, was also biased in the initial PEE.

Optimised perturbed emissions ensemble

To reduce the high bias in the initial PEE, the lower bounds of uncertainty of most parameters were decreased (Table 1, column *Optimised PEE*). The upper bound of uncertainty of transport emissions was also reduced. For other parameters, the elicited upper bounds were retained due to lack of evidence supporting a decrease. The uncertainty range of most parameters was therefore widened.

Power sector emissions were separated into emissions below and above 152 m based on the vertical distribution of the base emissions, which peak in the 4th layer (top-of-layer at 152 m). Perturbations to the vertical profile were thus enabled. The absolute maximum value from expert elicitation was set as the upper bound and a newly defined lower bound was used, allowing for a wide uncertainty range.

We also modified the definition of night-time transport emissions as those occurring between 0 am and 5 am (inclusive), a time frame in which there were no restrictions for any types of vehicles in 2016 (Zhang et al. 2019). This fraction represents 9% of the total traffic sources in the base emissions. As the base run generally overestimated daytime concentrations while underestimating night-time levels (see Fig. S2), we used an uncertainty range of 10–30%.

The number of emission parameters was thus increased to 14. We drew 140 samples from the new 14-dimensional parameter space to construct an optimised PEE. Two examples (with the NO_x emissions parameters omitted) are shown in Fig. 3b, while Fig. 3a summarises the key steps in the methodology presented in this section. The annual total CO emissions and their vertical and mean diurnal distributions in each optimised PEE member, compared to those in the base emissions, are shown in Fig. 4. The optimised PEE was then used in simulations forced with the same meteorological data and background concentrations from 2016 as the initial PEE simulations.

Results and discussion

As with the evaluation of the initial PEE simulations (see Section ‘Simulation setup and initial evaluation’), we compared local CO concentrations simulated with the optimised PEE against those observed at the long-term monitoring sites in 2016. The mean square error (MSE) in hourly local CO concentrations was calculated for each simulation at each site. The MSE is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (mod_i - obs_i)^2 \quad (2)$$

where obs_i is the local CO concentration observed in hour i , mod_i is the corresponding simulation output and n is the length of the hourly observations available. Following Solazzo and Galmarini (2016), MSE can be decomposed as the sum of three components:

$$MSE = \left(\overline{mod} - \overline{obs} \right)^2 + (\sigma_{mod-r\sigma_{obs}})^2 + \sigma_{obs}^2 \times (1-r^2) \quad (3)$$

where r is the Pearson’s correlation coefficient between the modelled and observed concentrations.

The first two terms in Eq. (3) represent errors associated with the bias and the variance, which measure the accuracy and the precision of simulations, respectively. The last term reflects the fraction of the observed variance not explained by the model and is referred to as the minimum achievable MSE (mMSE). As explained in Section ‘Simulation setup and initial evaluation’, the NMBF gives the magnitude of the factor by which simulation outputs differ from observations and its sense (i.e. over- or underestimation) (Yu et al. 2006), the latter was essential in adjusting the elicited uncertainty ranges. Yet NMBF only measures model errors associated with the bias, whereas the MSE is a more compact metric that estimates both systematic and random errors (Solazzo and Galmarini 2016). The benefits of using the MSE for model performance evaluation are demonstrated in the following.

As shown in Fig. 5a, MSEs of the optimised PEE simulations span a wide range at most urban and traffic monitoring sites, while other sites are mainly associated a narrower range of MSEs. This indicates higher base emissions and thus larger variations between PEE members in the central urban areas than the periphery with the same perturbations. The sites LLH, YLD, YF (regional background sites) and DX (a suburban site), all located in the south of the modelling domain, are associated with the largest MSEs. This is evident from the minimum MSEs (i.e. the lower end of the whiskers) which are the largest at these sites. MSE breakdown in Fig. 5b shows that the variance and bias errors are larger than the mMSEs at most sites. The MSEs at these four sites, however, are dominated by the mMSEs, which can be further decomposed

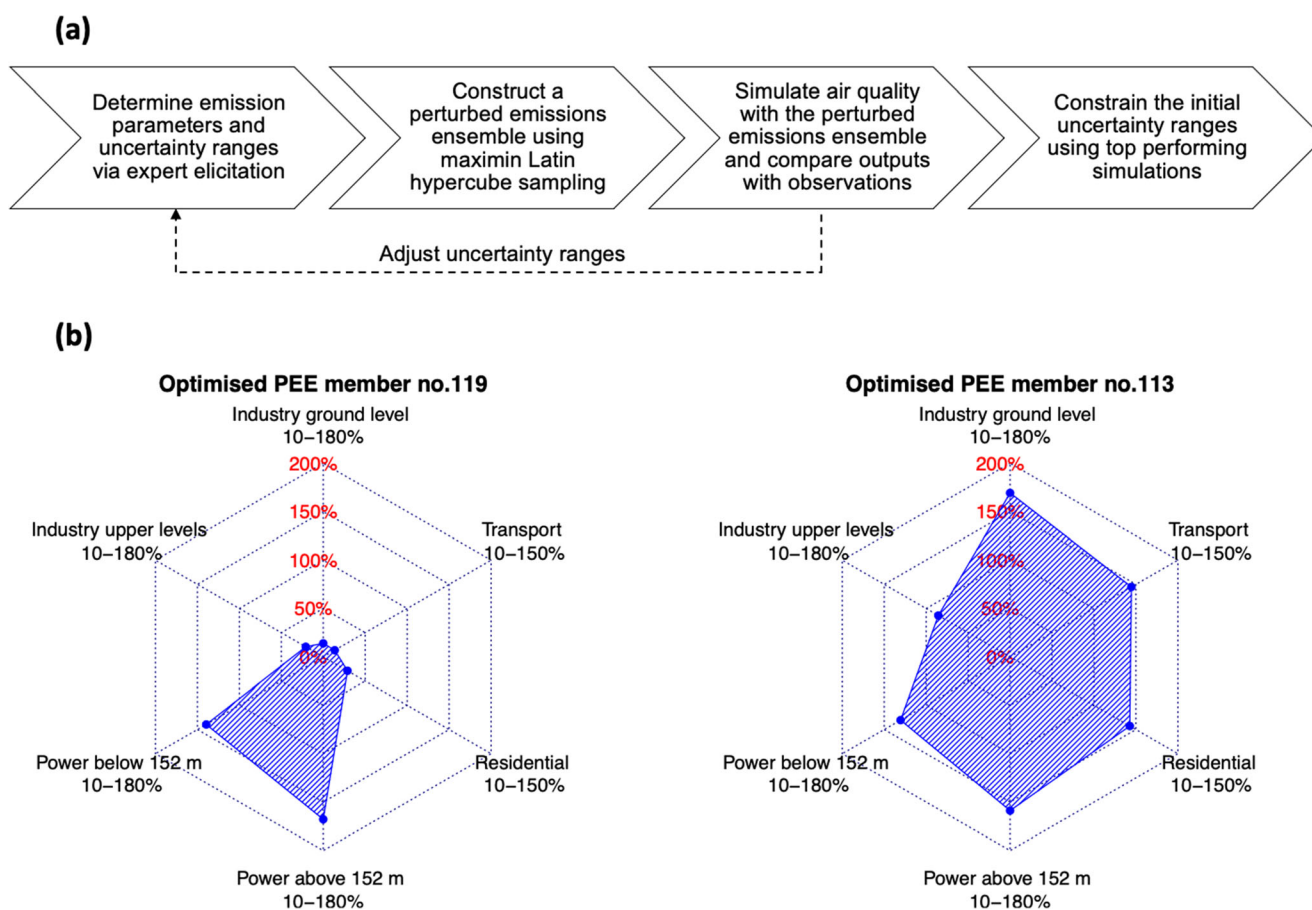


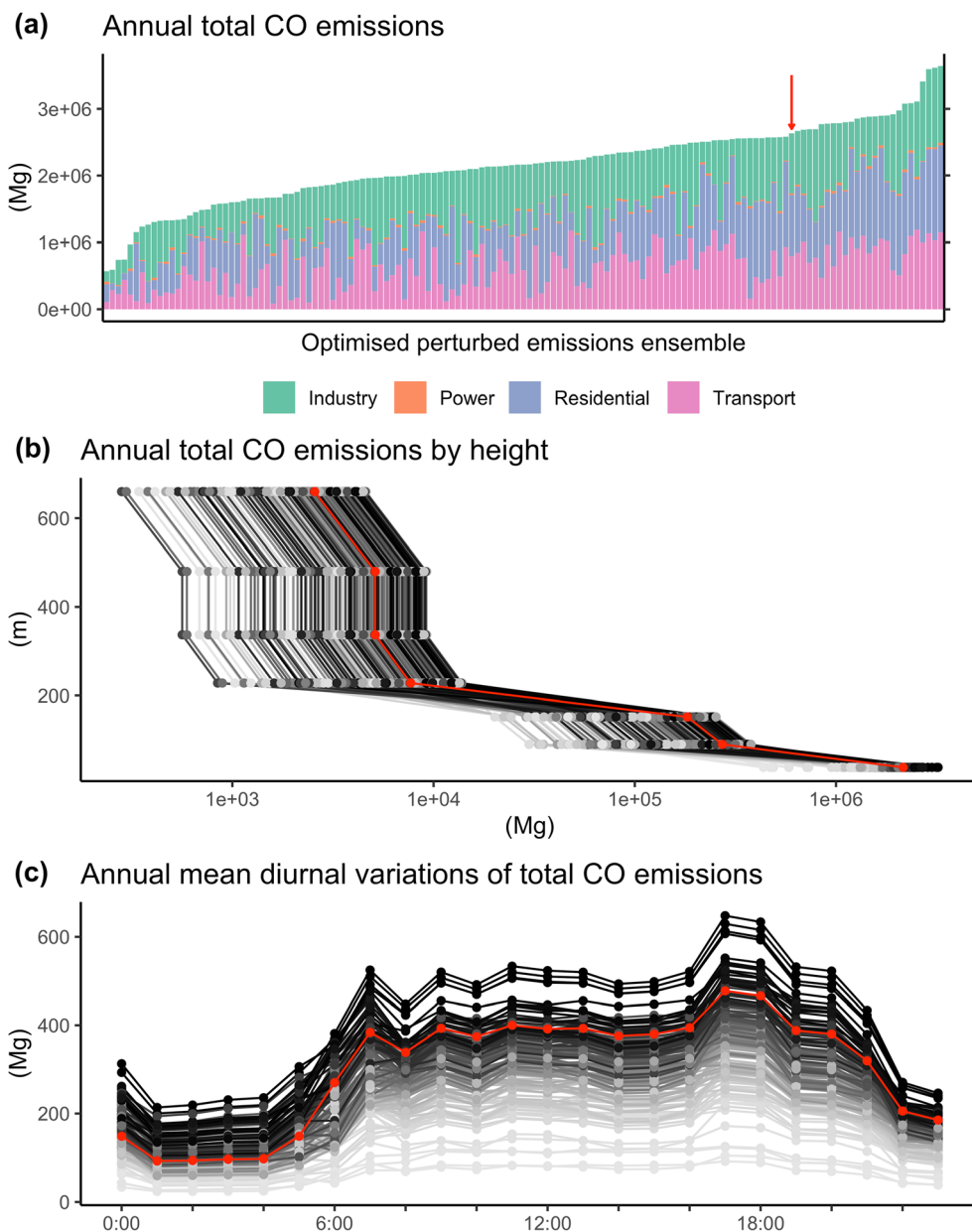
Fig. 3 (a) Main steps of the methodology presented in this paper. (b) Design of the optimised perturbed emissions ensemble (PEE) with two members as examples. The NO_x emissions parameters are omitted as they are beyond the scope of this work. Each axis of the hexagon represents a CO emissions parameter, the uncertainty range of which is labelled in black. Note that the uncertainty ranges vary between parameters,

but the same scale is used on all axes with labels shown on one (red). Blue dots show the values set for individual parameters in a specific member. The shaded area connecting the dots represents how the multidimensional uncertainty space is sampled by that member. The parameter for the night-time fraction of transport emissions is also not shown, as its uncertainty range substantially differs from those shown here

according to Eq. (3). As the correlation coefficients are low at all sites due to removal of background concentrations (Fig. 5c), the mMSEs are mostly made up of the variance in observations. The four sites are in fact associated with the highest observed variance, mainly from the winter months (Fig. 5d). The overall higher variance in CO concentrations in winter results from the frequent occurrence of severe pollution episodes in the study area due to enhanced emissions and stagnant weather conditions (Chen and Wang 2015). The particularly high variance at the aforementioned sites can be caused by local concentrations well above network averages during such episodes. Hua et al. (2018) found that the much higher CO, NO_2 and $\text{PM}_{2.5}$ concentrations compared to the city-wide means at the site LLH in winter 2015/2016 could be attributed to local coal combustion and biomass burning for residential heating, while natural gas powered central heating (with much lower emissions) was provided in urban Beijing. The sites YLD, YF and DX were likely affected by similar emissions.

Each PEE simulation is associated with 33 MSEs (i.e. one for each monitoring site), the distribution of which is skewed due to high values as discussed above. Therefore, we used the median MSE to represent a simulation's average performance across the modelling domain (Fig. 6), instead of the mean MSE which is inflated by the extreme values. The UK Department for Environment, Food and Rural Affairs also uses the median to represent network average concentrations in their 'Air Pollution in the UK' reports. It is more robust than the arithmetic mean particularly when a site with extreme concentrations starts or ceases operating (Department for Environment Food and Rural Affairs 2020). Different combinations of emission parameters may result in the same concentration at a particular location. By using observations from a site network that samples a wide range of emission strengths and source mixes, the risk of fitting the PEE to just one of several possible parameter settings can be reduced. Figure 6 reveals that in the top performing optimised PEE simulations, the bias and variance components are small and the remaining

Fig. 4 (a) Annual total CO emissions and contributions from individual source sectors in the optimised perturbed emissions ensemble (PEE) and the base emissions (marked by the red arrow). (b) Vertical distribution of the annual total CO emissions in the optimised PEE and the base emissions. The height is given as the top of layer height above ground. (c) Annual mean diurnal variations in total CO emissions in the optimised PEE and the base emissions. In panels (b) and (c), the base emissions are coloured in red and the optimised PEE members are colour-coded according to their total CO emissions with darker colour indicating higher values



MSEs are mostly made up of the mMSE. As the bias is mainly introduced by external forcings, it is greatly reduced with improved emissions estimates. The variance error and the mMSE are associated with internal processes in the model; thus, they are more relevant from a model development viewpoint, less so in the context of model application (as is the purpose of constraining emission estimates). The mMSE is the least troublesome amongst the error components (Solazzo and Galmarini 2016), because it reflects non-systematic errors including noise in the observations as discussed above.

We also calculated each optimised PEE simulation’s MSE in the annual mean diurnal variations of local CO concentrations at each monitoring site. This was to constrain the uncertainty range of the night-time fraction of traffic emissions

which affects the diurnal distribution of concentrations. The median MSE associated with a simulation was again used as a measure of its overall skill in simulating the diverse diurnal profiles within the modelling domain.

Figure 7 shows the average performance of individual optimised PEE simulations against the value set for each emission parameter in Table 1. It is evident from Fig. 7f that transport emission estimates are substantially lower in the top performing simulations (as measured by the median MSE in hourly local CO concentrations) than those in the base run. In simulations with a median MSE within the 1st quartile, transport emissions are 11–86% of those in the base emissions. This range reduces with decreasing median MSE, such that in simulations with a median MSE below the 5th percentile,

Fig. 5 Distribution of (a) the mean square errors, (b) the components of mean square errors and (c) the Pearson’s correlation coefficients in hourly local (i.e. background removed) CO concentrations calculated for the optimised perturbed emissions ensemble (PEE) simulations (shown as dim grey box-and-whisker plots) and the base run (red line segments) at each long-term monitoring site. (d) Annual and seasonal variance in the observed hourly local CO concentrations at each site. Winter (DJF) variances are represented by the y-axis on the right, while variances in other seasons and the annual variances are represented by the y-axis on the left. In all panels, the monitoring sites are colour-coded according to the site type: urban site (magenta), traffic monitoring site (purple), suburban site (orange), clean site (light green) and regional background site (green)

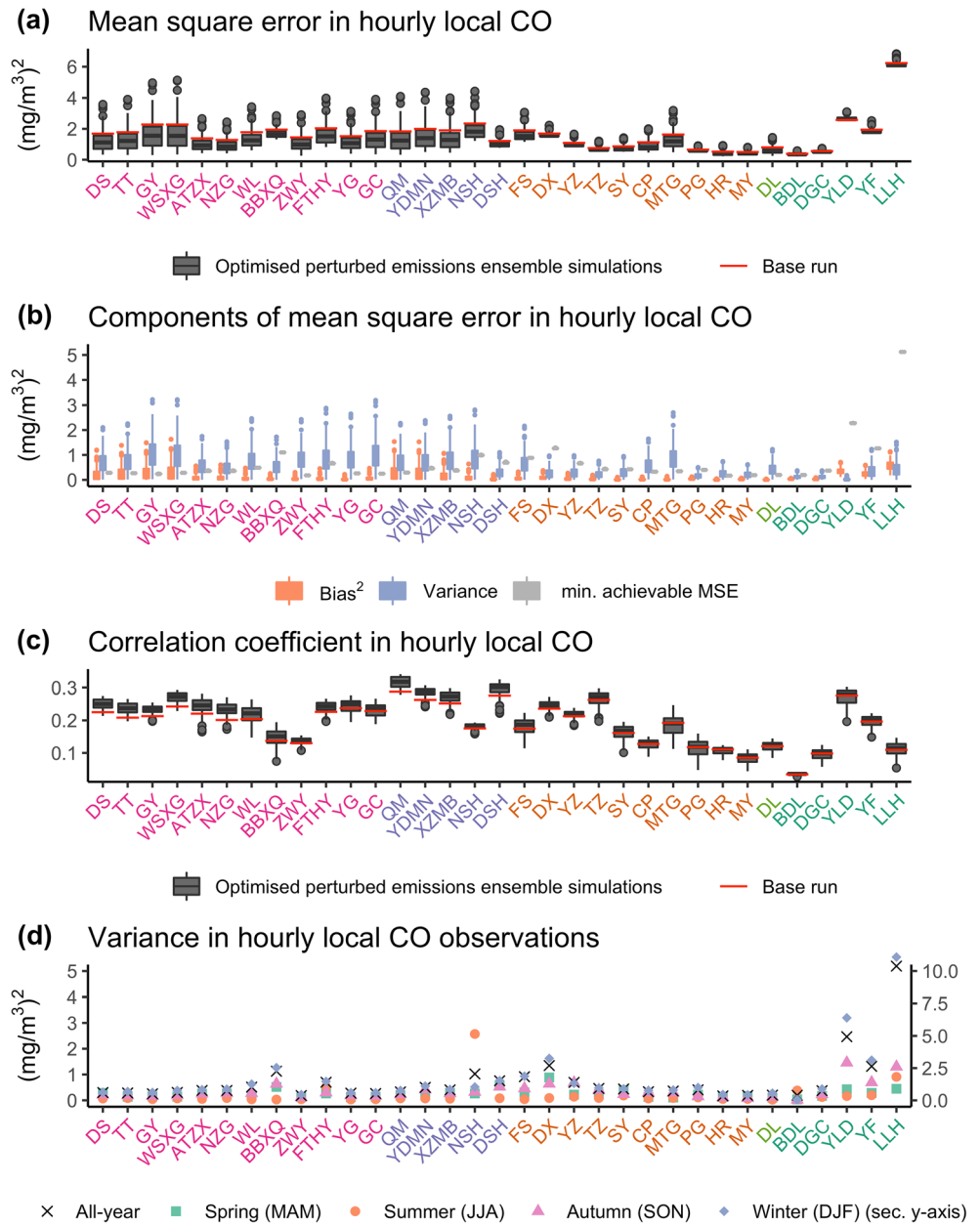


Fig. 6 Median mean square errors in hourly local (i.e. background removed) CO concentrations of the optimised perturbed emissions ensemble (PEE) simulations and the base run (marked by the red arrow), sorted in ascending order and decomposed into bias, variance and minimum achievable mean square error

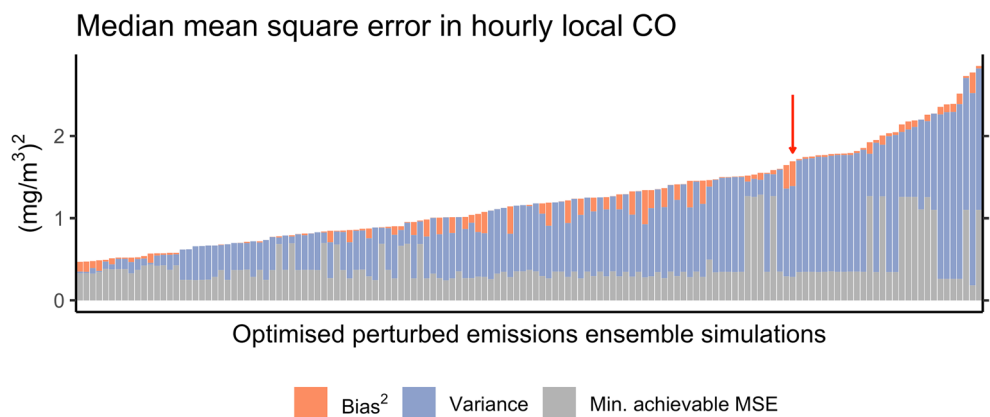
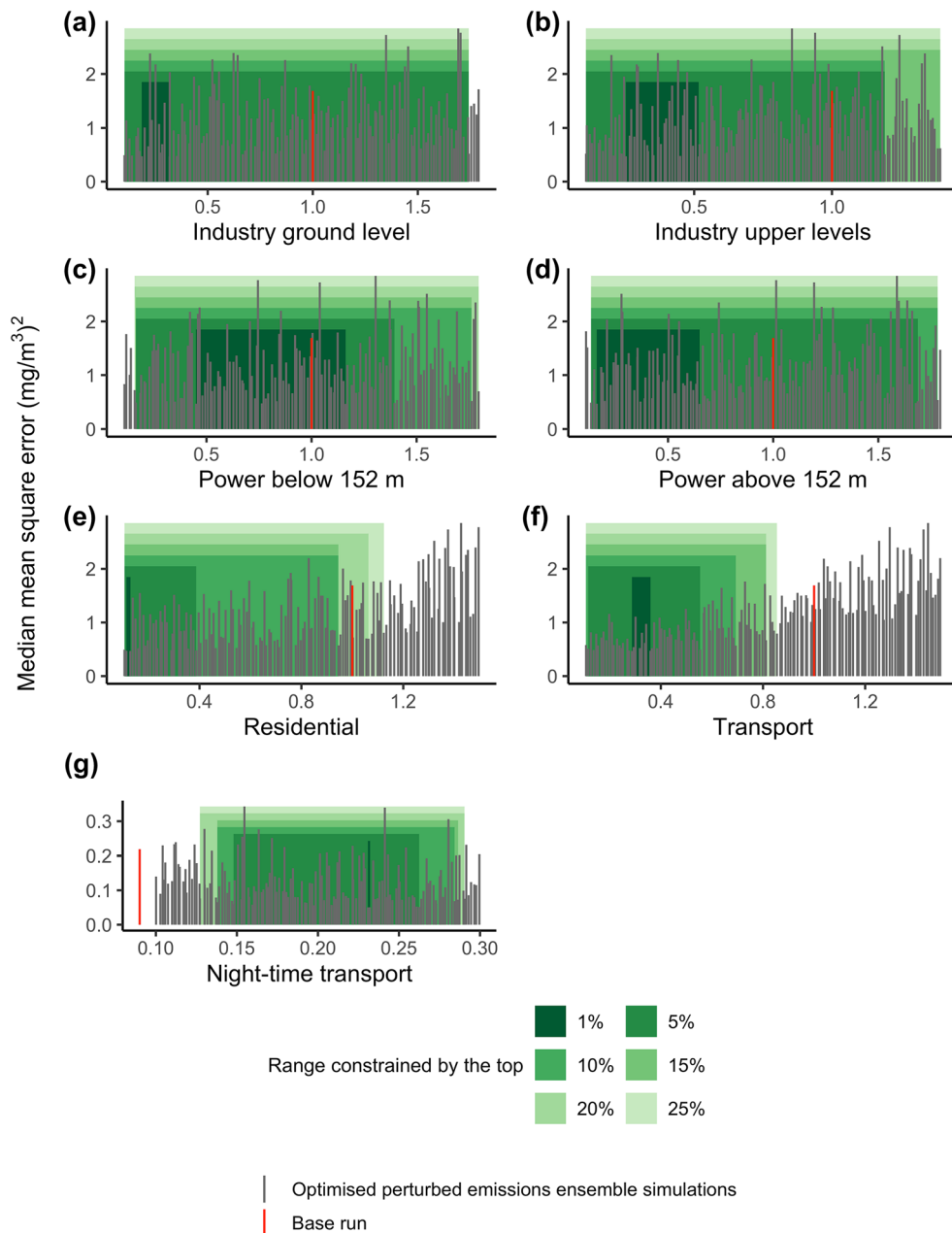


Fig. 7 Performance of the optimised perturbed emissions ensemble (PEE) simulations (dim grey bars) and the base run (red bars) as a function of emission parameter values. The scales on the x-axes correspond to the uncertainty ranges in Table 1. Green shades mark the range of parameter values in the top performing 1%, 5%, 10%, 15%, 20% and 25% of the simulations, as measured by their median mean square errors in hourly local (i.e. background removed) CO concentrations at the long-term monitoring sites across the modelling domain (see Fig. 6) in all subplots except in (g), where median mean square errors in the mean diurnal variations of local CO concentrations are used (note the different scale on the y-axis)



traffic sources are reduced to 12–56% of those in the base emissions. As discussed earlier, this predominantly reflects the trends in emissions between 2013 (the year for which the inventory was compiled) and 2016 (when the observations were made). It is also worth noting that 1% of a 140-member ensemble only consists of two members. As shown in Fig. 6, differences in the median MSEs of the top performing simulations are small; thus, the parameter ranges constrained by the top 1% were regarded as not robust. In addition to the total traffic emissions, Fig. 7g reveals that the uncertainty range for the night-time proportion can also be constrained. Amongst the top 5% of the simulations for modelling the mean diurnal cycles, the night-time fraction of

traffic emissions varies between 16% and 26%, substantially higher than the 9% in the base emissions. Unlike the transport sector, though the maximum parameter value for residential emissions (Fig. 7e) also decreases with improving simulation performance, the minimum value is fixed at its lower bound of uncertainty (see Table 1) except in the top performing 1%. In other words, the uncertainty range can only be constrained on one side — residential emissions are likely to be below 39% of those in the base emissions, if the top performing 5% of the simulations are considered. In terms of industry and power sector emissions (Fig. 7a–d), however, values in the top performing 25% of the simulations often span the entire

uncertainty range of the individual parameters. The ranges only start to decrease when the median MSE falls below the 10th percentile. Even ranges in the top 1% are wider than those for the transport and residential sectors. Hence, the current PEE and existing observations do not provide enough constraints for these emission parameters.

Figure S3 reveals that the contribution of power sector emissions to the annual mean local CO concentration is minimal at all long-term monitoring sites in the base run, as power sources are negligible in the total base emissions (see Fig. 4a). An increase by 80% (i.e. upper bound of uncertainty, see Table 1) is too small to be reflected in changes in the simulated concentrations, particularly as the output locations are not within the influence region of any power sources, which are only present in a few grid cells. This can also be seen in the source apportionment by grid cell for the site TZ (Fig. S4).

There are several possible explanations for the relatively small contributions to the CO concentrations at most sites from industrial emissions (Fig. S3), which makes up approximately 1/3 of the total base emissions (see Fig. 4a). Industrial sources may be under-constrained as emissions were split into two parameters and perturbed simultaneously (i.e. not in a correlated manner) for varying vertical profiles (see Table 1). The perturbed range of the sum of the two parameters, the total industrial emissions, thus may not be large enough to capture the real emissions. Additionally, the existing ground-based observations may not provide sufficient constraints for the vertically distributed emissions. Moreover, it is evident from Fig. S4 that most industrial sources are concentrated in the east and south of the modelling domain; thus, their influence on the concentrations at most monitoring sites may have a stronger dependence on the wind direction, compared to the more ubiquitous residential and transport sources. To test this hypothesis, we further evaluated the performance of the optimised PEE simulations in each season. As can be seen from Fig. 8, the ranges in industrial ground and elevated CO emissions reduce with decreasing median MSE in hourly local CO concentrations in spring and summer, when the winds are mainly from the south. In autumn and winter with prevailing northerly winds (Zhao et al. 2011), the ranges are only reduced when the median MSE falls below the 5th percentile. This indicates that industrial emissions have a larger influence on the CO concentrations at the monitoring sites when located in the upwind direction. However, unlike the residential and transport sectors (see Fig. 7e and f), no conclusion can be drawn with regard to the sector total emissions in 2016.

Finally, we also evaluated the performance of the optimised PEE in simulating local CO concentrations measured by SNAQ boxes during the APHH-Beijing winter campaign (Fig. S5). Using this short-term, independent set of observations and the evaluation method described above, we were again able to constrain the uncertainty range of the

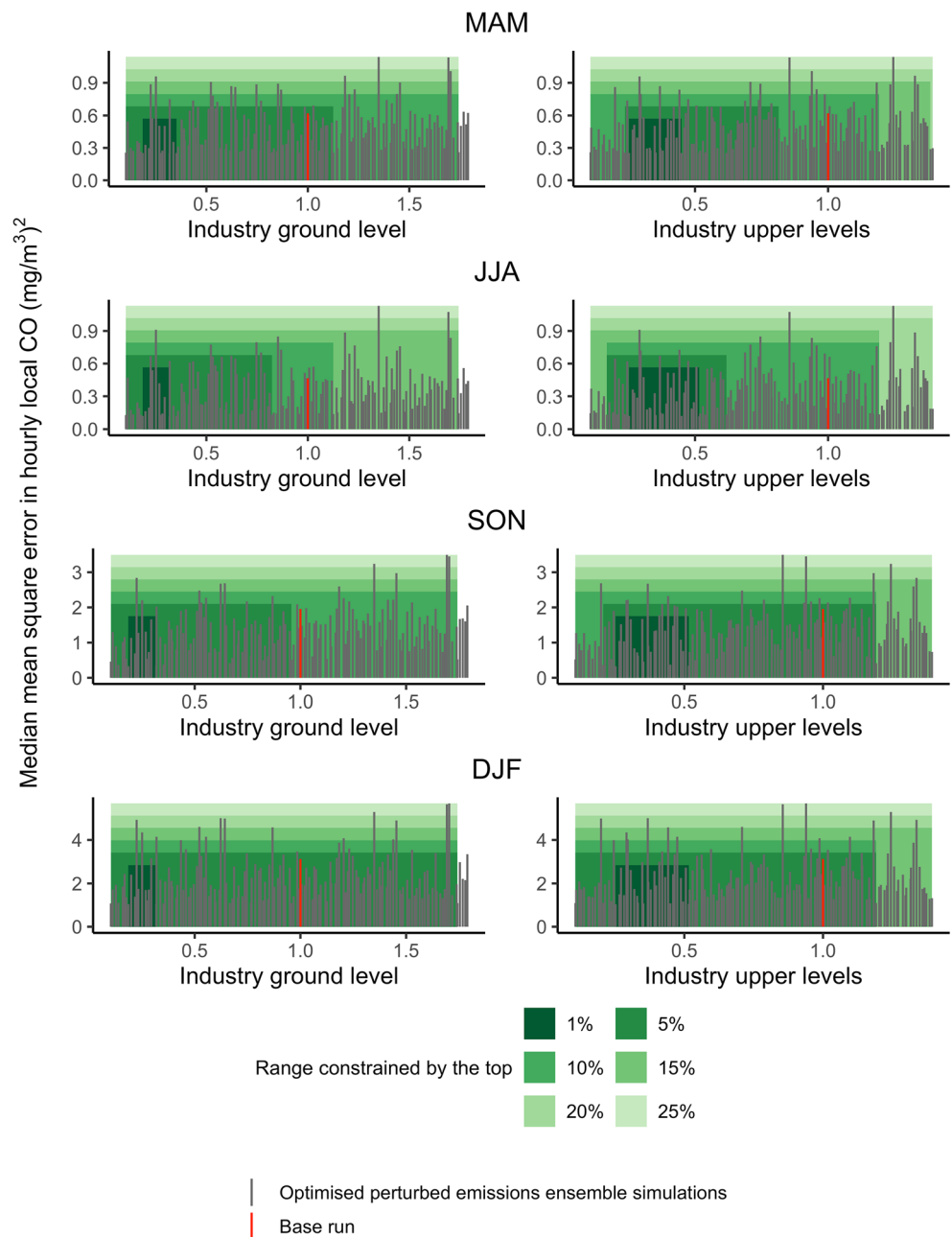
transport sources and (partly) that of the residential emissions (Fig. S6). Uncertainty ranges of the traffic and residential sources in the top 5% of the simulations are identical to those found using the full year's reference measurements as constraints, i.e. 12–56% and <39% of the base emissions, respectively, providing further support for the robustness of the results.

Discussion and conclusion

Annual CO emissions from the industry, power, transport and residential sectors are 9×10^5 , 2.6×10^4 , 7.9×10^5 and 9.1×10^5 Mg, respectively, in the base emissions from 2013 for the study area covering most of Beijing and parts of Hebei Province. In 2016, transport and residential sources likely decreased to $0.92\text{--}4.4 \times 10^5$ Mg and under 3.5×10^5 Mg (i.e. 44–88% and >61% reductions from 2013), respectively, based on the uncertainty ranges constrained by the top performing 5% of the optimised PEE simulations. These emission reductions provide strong and independent evidence for the effectiveness of relevant pollution control measures implemented in the study area. Annual industrial and power sector emissions in 2016 could not be constrained with the current PEE and existing observations. Hence, though our analysis suggested a downward trend in the total anthropogenic CO emissions within the study area from 2013 to 2016 (see Fig. 2), the magnitude of which could not be quantified.

Squires et al. (2020) compared the same base emissions with pollutant fluxes measured from a tower in central Beijing. On average, 90% of the flux was contributed by a footprint within 2 km from the tower, but contributions from up to 7 km away were measured depending on wind and atmospheric stability. They found that using CO base emissions from 2013 overestimated the observed fluxes by a mean factor of 4.8 during the APHH-Beijing winter campaign in 2016 and a mean factor of 10 during the summer campaign in 2017. Reducing the base emissions by 30% in winter and 43% in summer, the modelled fluxes were still on average 3.4 and 5.6 times higher in the respective seasons. They further considered only the traffic and residential sectors, as no industrial or power sources were identified within the footprint, and overestimations with a factor of 3.1 in winter and 5.2 in summer were still identified. Our findings suggest the sum of traffic and residential sources was likely below 7.9×10^5 Mg (taking the upper bound of uncertainty for both sectors) in 2016. The corresponding value in the 2013 base emissions is thus at least 2.2 times higher. This result has been derived for the entire modelling domain with no spatiotemporally varying changes assumed. It is broadly consistent with the factor of 3.1 found for the much smaller flux footprint during the APHH-Beijing winter campaign by Squires et al. (2020). They also revealed an underestimated night-time fraction in the diurnal

Fig. 8 Seasonal performance of the optimised perturbed emissions ensemble (PEE) simulations (dim grey bars) and the base run (red bars) as a function of the values set for the parameter industry sector ground level and elevated CO emissions. The scales on the x-axes correspond to the uncertainty ranges in Table 1. Green shades mark the range of parameter values in the top performing 1%, 5%, 10%, 15%, 20% and 25% of the simulations, as measured by their median mean square errors in hourly local (i.e. background removed) CO concentrations at long-term monitoring sites across the modelling domain, calculated for each season (note the different scales on the y-axes in different rows)



profile of CO base emissions (as the sum of the transport and residential sources) compared to flux measurements.

The standard MEIC v1.3 provides provincial-level totals from which the base emissions were downscaled and gridded (Zheng et al. 2017). The open-access version includes emission estimates for 2012, 2014 and 2016 (<http://www.meicmodel.org>, last accessed 22 August 2020). The emission reductions are smaller than those in our findings. In Beijing, transport emissions in 2016 were 63% and 91% of those in 2012 and 2014, respectively. Residential sources in 2016 were 71% and 86% of those in 2012 and 2014. Even smaller reductions were calculated for Hebei. Another bottom-up inventory for Beijing (Xue et al. 2019) revealed a

40% reduction in total CO emissions from 2013 to 2015. Different source categories were used such that there was no residential sector comparable to that in the MEIC, but a 43% decrease in mobile sources (including on-road and non-road vehicles) was found, more consistent with the 44–88% decrease in transport emissions between 2013 and 2016 found in this study. The China Vehicle Environmental Management Annual Reports calculated 26% and 0.4% reduction in on-road vehicle emissions from 2013 to 2016 for Beijing and Hebei, respectively. Conversely, the vehicle population grew by 6% and 78%, reflecting the effects of tougher emission standards, cleaner fuel standards and fleet turnover (Ministry of Environmental Protection of the People's Republic of

China 2014, 2017). Yet no exact comparisons can be made between these figures and our estimates due to the different spatial extents and time frames of the inventories used.

We were unable to constrain the uncertainties in the annual industrial and power sector CO emissions in 2016. This can partly be attributed to a lack of adequate observations, particularly those within the region of influence of large point sources, while the more diffuse traffic and residential sources are better sampled by the current ground-based monitoring network. We have also demonstrated that parameters can be under-constrained when their uncertainty ranges are poorly defined, highlighting the importance of a rigorous expert elicitation. Lastly, our findings are subject to several sources of uncertainty. We have not quantified the inherent uncertainties in the base emissions which may arise from inadequate emission factors or energy statistics, yet these are most likely small compared to the uncertainties associated with emission trends (see Section ‘Expert elicitation of emissions uncertainties’). Though we have not directly assessed the uncertainty in the input meteorology, we have demonstrated that the modelled local CO concentrations are much more sensitive to variations in the input emissions than the inter-annual variability in meteorology (Fig. S7). The effect of any uncertainties in the observations on the results is also expected to be small, as suggested by the identical results obtained using an independent set of observations. Overall, we believe this method of combining PEEs with observations (including short-term, low-cost sensor measurements) is widely applicable in providing timely updates of emission estimates by source sector for regions undergoing rapid changes and testing the effectiveness of the implemented emission reduction measures.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11869-021-01041-7>.

Acknowledgements We thank Oliver Wild, Michael Holloway and Michael Biggart for processing the base emissions. Thanks also go to Sue Grimmond and Simone Kotthaus for providing mixed layer height measurements and comments on the manuscript.

Data and code availability The Multi-resolution Emission Inventory for China (MEIC) is available upon request at <http://www.meicmodel.org>. Long-term air quality monitoring data from Beijing is archived at <https://quoteft.net/air/> and <http://data.epmap.org>. Licences for ADMS-Urban can be purchased at <https://www.cerc.co.uk/>. Codes used to generate the perturbed emissions ensemble using the R language are available at: <https://github.com/yuanle731/PEE>. We believe that the methodology presented is generalisable to other air quality models, emissions inventories and air quality measurements from other regions.

Funding This work was supported by the Natural Environment Research Council (NE/N007093/1), the National Natural Science Foundation of China (No. 42061130213), the Royal Society of UK through Newton Advanced Fellowship (NAFR1/201166) and the Tsinghua University Initiative Scientific Research Programme. Alexander T. Archibald acknowledges National Centre for Atmospheric Science and the Met Office through the Clean Air SPF for funding.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arya SP (1999) Air pollution meteorology and dispersion. Oxford University Press, New York
- Azzi M, Johnson GM, Cope M (1992) An introduction to the generic reaction set photochemical smog mechanism. In: Proceedings of the 11th International Conference of the Clean Air Society of Australia and New Zealand. 451–462
- Baldocchi DD (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Glob Chang Biol* 9:479–492. <https://doi.org/10.1046/j.1365-2486.2003.00629.x>
- Biggart M, Stocker J, Doherty RM, Wild O, Holloway M, Carruthers D, Li J, Zhang Q, Wu R, Kotthaus S, Grimmond S, Squires FA, Lee J, Shi Z (2020) Street-scale air quality modelling for Beijing during a winter 2016 measurement campaign. *Atmos Chem Phys* 20:2755–2780. <https://doi.org/10.5194/acp-20-2755-2020>
- Brasseur GP, Xie Y, Petersen AK, Bouarar I, Flemming J, Gauss M, Jiang F, Kouznetsov R, Kranenburg R, Mijling B, Peuch VH, Pommier M, Segers A, Sofiev M, Timmermans R, van der A R, Walters S, Xu J, Zhou G (2019) Ensemble forecasts of air quality in eastern China – Part 1: Model description and implementation of the MarcoPolo–Panda prediction system, version 1. *Geosci Model Dev* 12:33–67. <https://doi.org/10.5194/gmd-12-33-2019>
- Brierley CM, Collins M, Thorpe AJ (2010) The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. *Clim Dyn* 34:325–343. <https://doi.org/10.1007/s00382-008-0486-3>
- Cambridge Environmental Research Consults Limited (2017) ADMS-Urban Urban Air Quality Management System Version 4.1 user guide. 376
- Carruthers D, Stidworthy A, Clarke D, Dicks J, Jones R, Leslie I, Popoola OAM, Seaton M (2019) Urban emission inventory optimisation using sensor data, an urban air quality model and inversion techniques. *Int J Environ Pollut* 66:252–266. <https://doi.org/10.1504/IJEP.2019.104878>
- Chen H, Wang H (2015) Haze Days in North China and the associated atmospheric circulations based on daily visibility data from 1960 to 2012. *J Geophys Res Atmos* 120:5895–5909. <https://doi.org/10.1002/2015JD023225>
- Chen W, Tang H, Zhao H (2015) Diurnal, weekly and monthly spatial variations of air pollutants and air quality of Beijing. *Atmos Environ* 119:21–34. <https://doi.org/10.1016/j.atmosenv.2015.08.040>
- Cheng J, Su J, Cui T, Li X, Dong X, Sun F, Yang Y, Tong D, Zheng Y, Li Y, Li J, Zhang Q, He K (2019) Dominant role of emission reduction

- in PM_{2.5} air quality improvement in Beijing during 2013–2017: a model-based decomposition analysis. *Atmos Chem Phys* 19:6125–6146. <https://doi.org/10.5194/acp-19-6125-2019>
- Cheng J, Tong D, Liu Y, Bo Y, Zheng B, Geng G, He K, Zhang Q (2020) Air quality and health benefits from China's current and upcoming clean air policies. *Faraday Discuss* 226:584–606. <https://doi.org/10.1039/D0FD00090F>
- Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, Balakrishnan K, Brunekreef B, Dandona L, Dandona R, Feigin V, Freedman G, Hubbell B, Jobling A, Kan H, Knibbs L, Liu Y, Martin R, Morawska L, Pope CA III, Shin H, Straif K, Shaddick G, Thomas M, van Dingenen R, van Donkelaar A, Vos T, Murray CJL, Forouzanfar MH (2017) Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* 389:1907–1918. [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6)
- Colette A, Andersson C, Manders A, Mar K, Mircea M, Pay MT, Raffort V, Tsyro S, Cuvelier C, Adani M, Bessagnet B, Bergström R, Briganti G, Butler T, Cappelletti A, Couvidat F, D'Isidoro M, Doumbia T, Fagerli H, Granier C, Heyes C, Klimont Z, Ojha N, Otero N, Schaap M, Sindelarova K, Stegehuis AI, Roustan Y, Vautard R, van Meijgaard E, Vivanco MG, Wind P (2017) EURODELTA-Trends, a multi-model experiment of air quality hindcast in Europe over 1990–2010. *Geosci Model Dev* 10:3255–3276. <https://doi.org/10.5194/gmd-10-3255-2017>
- Collins M, Booth BBB, Bhaskaran B, Harris GR, Murphy JM, Sexton DMH, Webb MJ (2011) Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Clim Dyn* 36:1737–1766. <https://doi.org/10.1007/s00382-010-0808-0>
- Crippa M, Janssens-Maenhout G, Guizzardi D, van Dingenen R, Dentener F (2019) Contribution and uncertainty of sectorial and regional emissions to regional and global PM_{2.5} health impacts. *Atmos Chem Phys* 19:5165–5186. <https://doi.org/10.5194/acp-19-5165-2019>
- de Foy B, Wilkins JL, Lu Z, Streets DG, Duncan BN (2014) Model evaluation of methods for estimating surface emissions and chemical lifetimes from satellite data. *Atmos Environ* 98:66–77. <https://doi.org/10.1016/j.atmosenv.2014.08.051>
- Department for Environment Food & Rural Affairs (2020) Air pollution in the UK 2019. Department for Environment, Food and Rural Affairs (Defra), Nobel House, 17 Smith Square, London SW1P 3JR helpline@defra.gsi.gov.uk
- Geng G, Zhang Q, Martin RV, Lin J, Huo H, Zheng B, Wang S, He K (2017) Impact of spatial proxies on the representation of bottom-up emission inventories: a satellite-based analysis. *Atmos Chem Phys* 17:4131–4145. <https://doi.org/10.5194/acp-17-4131-2017>
- Granier C, Bessagnet B, Bond T, D'Angiola A, Denier van der Gon H, Frost GJ, Heil A, Kaiser JW, Kinne S, Klimont Z, Kloster S, Lamarque JF, Lioussé C, Masui T, Meleux F, Mieville A, Ohara T, Raut JC, Riahi K, Schultz MG, Smith SJ, Thompson A, van Aardenne J, van der Werf GR, van Vuuren DP (2011) Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Clim Chang* 109:163–190. <https://doi.org/10.1007/s10584-011-0154-1>
- He B, Heal MR, Humstad KH, Yan L, Zhang Q, Reis S (2019) A hybrid model approach for estimating health burden from NO₂ in megacities in China: a case study in Guangzhou. *Environ Res Lett* 14:124019. <https://doi.org/10.1088/1748-9326/ab4f96>
- Hoesly RM, Smith SJ, Feng L, Klimont Z, Janssens-Maenhout G, Pitkanen T, Seibert JJ, Vu L, Andres RJ, Bolt RM, Bond TC, Dawidowski L, Kholod N, Kurokawa JI, Li M, Liu L, Lu Z, Moura MCP, O'Rourke PR, Zhang Q (2018) Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geosci Model Dev* 11:369–408. <https://doi.org/10.5194/gmd-11-369-2018>
- Hong C, Zhang Q, He K, Guan D, Li M, Liu F, Zheng B (2017) Variations of China's emission estimates: response to uncertainties in energy statistics. *Atmos Chem Phys* 17:1227–1239. <https://doi.org/10.5194/acp-17-1227-2017>
- Hood C, MacKenzie I, Stocker J, Johnson K, Carruthers D, Vieno M, Doherty R (2018) Air quality simulations for London using a coupled regional-to-local modelling system. *Atmos Chem Phys* 18:11221–11245. <https://doi.org/10.5194/acp-18-11221-2018>
- Hua Y, Wang S, Jiang J, Zhou W, Xu Q, Li X, Liu B, Zhang D, Zheng M (2018) Characteristics and sources of aerosol pollution at a polluted rural site southwest in Beijing, China. *Sci Total Environ* 626:519–527. <https://doi.org/10.1016/j.scitotenv.2018.01.047>
- Janssens-Maenhout G, Crippa M, Guizzardi D, Dentener F, Muntean M, Pouliot G, Keating T, Zhang Q, Kurokawa J, Wankmüller R, Denier van der Gon H, Kuenen JJP, Klimont Z, Frost G, Darras S, Koffi B, Li M (2015) HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution. *Atmos Chem Phys* 15:11411–11432. <https://doi.org/10.5194/acp-15-11411-2015>
- Jiang Z, Worden JR, Worden H, Deeter M, Jones DBA, Arellano AF, Henze DK (2017) A 15-year record of CO emissions constrained by MOPITT CO observations. *Atmos Chem Phys* 17:4565–4583. <https://doi.org/10.5194/acp-17-4565-2017>
- Johnson ME, Moore LM, Ylvisaker D (1990) Minimax and maximin distance designs. *J Stat Plan Inference* 26:131–148. [https://doi.org/10.1016/0378-3758\(90\)90122-B](https://doi.org/10.1016/0378-3758(90)90122-B)
- Joseph VR, Hung Y (2008) Orthogonal-maximin Latin Hypercube Designs. *Stat Sin* 18:171–186
- Khalil MAK, Rasmussen RA (1990) The global cycle of carbon monoxide: trends and mass balance. *Chemosphere* 20:227–242. [https://doi.org/10.1016/0045-6535\(90\)90098-E](https://doi.org/10.1016/0045-6535(90)90098-E)
- Kotthaus S, Grimmond CSB (2018) Atmospheric boundary-layer characteristics from ceilometer measurements. Part 1: A new method to track mixed layer height and classify clouds. *Q J R Meteorol Soc* 144:1525–1538. <https://doi.org/10.1002/qj.3299>
- Lambert FH, Harris GR, Collins M, Murphy JM, Sexton DMH, Booth BBB (2013) Interactions between perturbations to different Earth system components simulated by a fully-coupled climate model. *Clim Dyn* 41:3055–3072. <https://doi.org/10.1007/s00382-012-1618-3>
- Lee LA, Pringle KJ, Reddington CL, Mann GW, Stier P, Spracklen DV, Pierce JR, Carslaw KS (2013) The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmos Chem Phys* 13:8879–8914. <https://doi.org/10.5194/acp-13-8879-2013>
- Lee JD, Helfter C, Purvis RM, Beevers SD, Carslaw DC, Lewis AC, Møller SJ, Tremper A, Vaughan A, Nemitz EG (2015) Measurement of NO_x fluxes from a tall tower in Central London, UK and comparison with emissions inventories. *Environ Sci Technol* 49:1025–1034. <https://doi.org/10.1021/es5049072>
- Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A (2015) The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* 525:367–371. <https://doi.org/10.1038/nature15371>
- Li Y, Ma Z, Han T, Quan W, Wang J, Zhou H, He D, Dong F (2020) Long-term declining in carbon monoxide (CO) at a rural site of Beijing during 2006–2018 implies the improved combustion efficiency and effective emission control. *J Environ Sci (China)*. <https://doi.org/10.1016/j.jes.2020.11.011>
- Liang C-K, West JJ, Silva RA, Bian H, Chin M, Davila Y, Dentener FJ, Emmons L, Flemming J, Folberth G, Henze D, Im U, Jonson JE, Keating TJ, Kucsera T, Lenzen A, Lin M, Lund MT, Pan X, Park RJ, Pierce RB, Sekiya T, Sudo K, Takemura T (2018) HTAP2 multi-model estimates of premature human mortality due to inter-continental transport of air pollution and emission sectors. *Atmos*

- Chem Phys 18:10497–10520. <https://doi.org/10.5194/acp-18-10497-2018>
- Liu F, Zhang Q, Tong D, Zheng B, Li M, Huo H, He KB (2015) High-resolution inventory of technologies, activities, and emissions of coal-fired power plants in China from 1990 to 2010. *Atmos Chem Phys* 15:13299–13317. <https://doi.org/10.5194/acp-15-13299-2015>
- Liu F, Zhang Q, Van Der ARJ et al (2016) Recent reduction in NO_x emissions over China: synthesis of satellite observations and emission inventories. *Environ Res Lett* 11. <https://doi.org/10.1088/1748-9326/11/11/114002>
- Loeppky JL, Sacks J, Welch WJ (2009) Choosing the sample size of a computer experiment: a practical guide. *Technometrics* 51:366–376. <https://doi.org/10.1198/TECH.2009.08040>
- Malkin TL, Heard DE, Hood C, Stocker J, Carruthers D, MacKenzie IA, Doherty RM, Vieno M, Lee J, Kleffmann J, Laufs S, Whalley LK (2016) Assessing chemistry schemes and constraints in air quality models used to predict ozone in London against the detailed Master Chemical Mechanism. *Faraday Discuss* 189:589–616. <https://doi.org/10.1039/c5fd00218d>
- McHugh CA, Carruthers DJ, Edmunds HA (1997) ADMS-Urban: an air quality management system for traffic, domestic and industrial pollution. *Int J Environ Pollut* 8:666–674. <https://doi.org/10.1504/IJEP.1997.028218>
- McKay MD, Beckman RJ, Conover WJ (1979) Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245. <https://doi.org/10.1080/00401706.1979.10489755>
- Ministry of Environmental Protection of the People's Republic of China (2013) HJ 654-2013. Specifications and test procedures for ambient air quality continuous automated monitoring system for SO₂, NO₂, O₃ and CO (in Chinese).
- Ministry of Environmental Protection of the People's Republic of China (2014) China Vehicle Environmental Management Annual Report 2014
- Ministry of Environmental Protection of the People's Republic of China (2017) China Vehicle Environmental Management Annual Report 2017
- Moore TO, Doughty DC, Marr LC (2009) Demonstration of a mobile Flux Laboratory for the Atmospheric Measurement of Emissions (FLAME) to assess emissions inventories. *J Environ Monit* 11: 259–268. <https://doi.org/10.1039/b810798j>
- Murphy JM, Sexton DMH, Barnett DH et al (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 430:768–772. <https://doi.org/10.1038/nature02771>
- Oakley JE, O'Hagan A (2019) SHELF: the Sheffield Elicitation Framework (version 4). School of Mathematics and Statistics, University of Sheffield, UK <http://tonyohagan.co.uk/shelf>
- Oda T, Maksyutov S (2011) A very high-resolution (1 km×1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmos Chem Phys* 11:543–556. <https://doi.org/10.5194/acp-11-543-2011>
- Owen B, Edmunds HA, Carruthers DJ, Singles RJ (2000) Prediction of total oxides of nitrogen and nitrogen dioxide concentrations in a large urban area using a new generation urban scale dispersion model with integral chemistry model. *Atmos Environ* 34:397–406. [https://doi.org/10.1016/S1352-2310\(99\)00332-5](https://doi.org/10.1016/S1352-2310(99)00332-5)
- Palmer TN (2000) Predicting uncertainty in forecasts of weather and climate. *Rep Prog Phys* 63:71–116. <https://doi.org/10.1088/0034-4885/63/2/201>
- Panagi M, Fleming ZL, Monks PS, Ashfold MJ, Wild O, Hollaway M, Zhang Q, Squires FA, Vande Hey JD (2020) Investigating the regional contributions to air pollution in Beijing: a dispersion modelling study using CO as a tracer. *Atmos Chem Phys* 20:2825–2838. <https://doi.org/10.5194/acp-20-2825-2020>
- Popoola OAM, Carruthers D, Lad C, Bright VB, Mead MI, Stettler MEJ, Saffell JR, Jones RL (2018) Use of networks of low cost air quality sensors to quantify air quality in urban settings. *Atmos Environ* 194: 58–70. <https://doi.org/10.1016/J.ATMOSENV.2018.09.030>
- Raupach MR, Rayner PJ, Paget M (2010) Regional variations in spatial structure of nightlights, population density and fossil-fuel CO₂ emissions. *Energy Policy* 38:4756–4764. <https://doi.org/10.1016/j.enpol.2009.08.021>
- Ruckstuhl AF, Henne S, Reimann S, Steinbacher M, Vollmer MK, O'Doherty S, Buchmann B, Hueglin C (2012) Robust extraction of baseline signal of atmospheric trace species using local regression. *Atmos Meas Tech* 5:2613–2624. <https://doi.org/10.5194/amt-5-2613-2012>
- Saide PE, Carmichael GR, Spak SN, Gallardo L, Osses AE, Mena-Carrasco MA, Pagowski M (2011) Forecasting urban PM₁₀ and PM_{2.5} pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model. *Atmos Environ* 45:2769–2780. <https://doi.org/10.1016/j.atmosenv.2011.02.001>
- Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis. *Environ Model Softw* 25:1508–1517. <https://doi.org/10.1016/j.envsoft.2010.04.012>
- Shi Z, Vu T, Kotthaus S, Harrison RM, Grimmond S, Yue S, Zhu T, Lee J, Han Y, Demuzere M, Dunmore RE, Ren L, Liu D, Wang Y, Wild O, Allan J, Acton WJ, Barlow J, Barratt B, Beddows D, Bloss WJ, Calzolari G, Carruthers D, Carslaw DC, Chan Q, Chatzidiakou L, Chen Y, Crilley L, Coe H, Dai T, Doherty R, Duan F, Fu P, Ge B, Ge M, Guan D, Hamilton JF, He K, Heal M, Heard D, Hewitt CN, Hollaway M, Hu M, Ji D, Jiang X, Jones R, Kalberer M, Kelly FJ, Kramer L, Langford B, Lin C, Lewis AC, Li J, Li W, Liu H, Liu J, Loh M, Lu K, Lucarelli F, Mann G, McFiggans G, Miller MR, Mills G, Monk P, Nemitz E, O'Connor F, Ouyang B, Palmer PI, Percival C, Popoola O, Reeves C, Rickard AR, Shao L, Shi G, Spracklen D, Stevenson D, Sun Y, Sun Z, Tao S, Tong S, Wang Q, Wang W, Wang X, Wang X, Wang Z, Wei L, Whalley L, Wu X, Wu Z, Xie P, Yang F, Zhang Q, Zhang Y, Zhang Y, Zheng M (2019) Introduction to the special issue “In-depth study of air pollution sources and processes within Beijing and its surrounding region (APHH-Beijing)”. *Atmos Chem Phys* 19:7519–7546. <https://doi.org/10.5194/acp-19-7519-2019>
- Silva RA, West JJ, Lamarque JF, Shindell DT, Collins WJ, Faluvegi G, Folberth GA, Horowitz LW, Nagashima T, Naik V, Rumbold ST, Sudo K, Takemura T, Bergmann D, Cameron-Smith P, Doherty RM, Josse B, MacKenzie IA, Stevenson DS, Zeng G (2017) Future global mortality from changes in air pollution attributable to climate change. *Nat Clim Chang* 7:647–651. <https://doi.org/10.1038/nclimate3354>
- Smith A, Lott N, Vose R et al (2011) The Integrated Surface Database: recent developments and partnerships. *Bull Am Meteorol Soc* 92: 704–708. <https://doi.org/10.1175/2011BAMS3015.1>
- Solazzo E, Galmarini S (2016) Error apportionment for atmospheric chemistry-transport models – a new approach to model evaluation. *Atmos Chem Phys* 16:6263–6283. <https://doi.org/10.5194/acp-16-6263-2016>
- Solazzo E, Riccio A, Van Dingenen R et al (2018) Evaluation and uncertainty estimation of the impact of air quality modelling on crop yields and premature deaths using a multi-model ensemble. *Sci Total Environ* 633:1437–1452. <https://doi.org/10.1016/j.scitotenv.2018.03.317>
- Squires FA, Nemitz E, Langford B, Wild O, Drysdale WS, Acton WJF, Fu P, Grimmond CSB, Hamilton JF, Hewitt CN, Hollaway M, Kotthaus S, Lee J, Metzger S, Pingingtha-Durden N, Shaw M, Vaughan AR, Wang X, Wu R, Zhang Q, Zhang Y (2020) Measurements of traffic-dominated pollutant emissions in a Chinese megacity. *Atmos Chem Phys* 20:8737–8761. <https://doi.org/10.5194/acp-20-8737-2020>
- Stewart ID, Oke TR (2012) Local climate zones for urban temperature studies. *Bull Am Meteorol Soc* 93:1879–1900. <https://doi.org/10.1175/BAMS-D-11-00019.1>

- Streets DG, Canty T, Carmichael GR, de Foy B, Dickerson RR, Duncan BN, Edwards DP, Haynes JA, Henze DK, Houyoux MR, Jacob DJ, Krotkov NA, Lamsal LN, Liu Y, Lu Z, Martin RV, Pfister GG, Pinder RW, Salawitch RJ, Wecht KJ (2013) Emissions estimation from satellite retrievals: a review of current capability. *Atmos Environ* 77:1011–1042
- Tai APK, Martin MV, Heald CL (2014) Threat to future global food security from climate change and ozone air pollution. *Nat Clim Chang* 4:817–821. <https://doi.org/10.1038/nclimate2317>
- Taksibi F, Khajepour H, Saboohi Y (2020) On the environmental effectiveness analysis of energy policies: a case study of air pollution in the megacity of Tehran. *Sci Total Environ* 705:135824. <https://doi.org/10.1016/j.scitotenv.2019.135824>
- Tang X, Zhu J, Wang ZF, Wang M, Gbaguidi A, Li J, Shao M, Tang GQ, Ji DS (2013) Inversion of CO emissions over Beijing and its surrounding areas with ensemble Kalman filter. *Atmos Environ* 81: 676–686. <https://doi.org/10.1016/j.atmosenv.2013.08.051>
- Vaughan AR, Lee JD, Misztal PK, Metzger S, Shaw MD, Lewis AC, Purvis RM, Carslaw DC, Goldstein AH, Hewitt CN, Davison B, Beevers SD, Karl TG (2016) Spatially resolved flux measurements of NO_x from London suggest significantly higher emissions than predicted by inventories. *Faraday Discuss* 189:455–472. <https://doi.org/10.1039/c5fd00170f>
- Venkatram A (1996) An examination of the Pasquill-Gifford-Turner dispersion scheme. *Atmos Environ* 30:1283–1290. [https://doi.org/10.1016/1352-2310\(95\)00367-3](https://doi.org/10.1016/1352-2310(95)00367-3)
- Wang S, Streets DG, Zhang Q, He K, Chen D, Kang S, Lu Z, Wang Y (2010) Satellite detection and model verification of NO_x emissions from power plants in Northern China. *Environ Res Lett* 5:044007. <https://doi.org/10.1088/1748-9326/5/4/044007>
- Wang S, Zhang Q, Martin RV, Philip S, Liu F, Li M, Jiang X, He K (2015) Satellite measurements oversee China's sulfur dioxide emission reductions from coal-fired power plants. *Environ Res Lett* 10: 114015. <https://doi.org/10.1088/1748-9326/10/11/114015>
- Xue Y, Zhang S, Zhou Z, Wang K, Liu K, Wang X, Shi A, Xu K, Tian H (2019) Spatio-temporal variations of multiple primary air pollutants emissions in Beijing of China, 2006–2015. *Atmosphere (Basel)* 10: 494. <https://doi.org/10.3390/atmos10090494>
- Yu S, Eder B, Dennis R, Chu SH, Schwartz SE (2006) New unbiased symmetric metrics for evaluation of air quality models. *Atmos Sci Lett* 7:26–34. <https://doi.org/10.1002/asl.125>
- Zhang S, Wu Y, Yan H, du X, Max Zhang K, Zheng X, Fu L, Hao J (2019) Black carbon pollution for a major road in Beijing: implications for policy interventions of the heavy-duty truck fleet. *Transp Res Part D Transp Environ* 68:110–121. <https://doi.org/10.1016/j.trd.2017.07.013>
- Zhao X, Zhang X, Pu W, Meng W, Xu X (2011) Scattering properties of the atmospheric aerosol in Beijing, China. *Atmos Res* 101:799–808. <https://doi.org/10.1016/j.atmosres.2011.05.010>
- Zhao Y, Nielsen CP, McElroy MB et al (2012) CO emissions in China: uncertainties and implications of improved energy efficiency and emission control. *Atmos Environ* 49:103–113. <https://doi.org/10.1016/j.atmosenv.2011.12.015>
- Zheng B, Zhang Q, Tong D, Chen C, Hong C, Li M, Geng G, Lei Y, Huo H, He K (2017) Resolution dependence of uncertainties in gridded emission inventories: a case study in Hebei, China. *Atmos Chem Phys* 17:921–933. <https://doi.org/10.5194/acp-17-921-2017>
- Zheng B, Tong D, Li M, Liu F, Hong C, Geng G, Li H, Li X, Peng L, Qi J, Yan L, Zhang Y, Zhao H, Zheng Y, He K, Zhang Q (2018) Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmos Chem Phys* 18:14095–14111. <https://doi.org/10.5194/acp-18-14095-2018>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.