Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale

B. Mijling and R. J. van der A

Received 22 March 2012; revised 12 June 2012; accepted 20 July 2012; published 5 September 2012.

Emission inventories of air pollutants are crucial information for policy makers and form important input data for air quality models. Using satellite observations for emission estimates has important advantages over bottom-up emission inventories: they are spatially consistent, have high temporal resolution, and enable updates shortly after the satellite data become available. We present a new algorithm specifically designed to use daily satellite observations of column concentrations for fast updates of emission estimates of short-lived atmospheric constituents on a mesoscopic scale ($\sim 25 \times 25$ km$^2$). The algorithm needs only one forward model run from a chemical transport model to calculate the sensitivity of concentration to emission, using trajectory analysis to account for transport away from the source. By using a Kalman filter in the inverse step, optimal use of the a priori knowledge and the newly observed data is made. We apply the algorithm for NO$_x$ emission estimates of East China, using the CHIMERE model on a 0.25 degree resolution together with tropospheric NO$_2$ column retrievals of the OMI and GOME-2 satellite instruments. Closed loop tests show that the algorithm is capable of reproducing new emission scenarios. Applied with real satellite data, the algorithm is able to detect emerging sources (e.g., new power plants), and improves emission information for areas where proxy data are not or badly known (e.g., shipping emissions). Chemical transport model runs with the daily updated emission estimates provide better spatial and temporal agreement between observed and simulated concentrations, facilitating improved air quality forecasts.


1. Introduction

Emission inventories are usually compiled from statistics on emitting activities and their typical emission factor. Despite the sustained efforts to provide complete and accurate databases, bottom-up emission inventories are not frequently updated, which limits their use. Changes related to recent economic developments, or specific political decisions toward emission reductions have a direct and measurable effect on atmospheric composition which is not accounted for by common emission inventories. As an example, Figure 1 shows the tropospheric NO$_2$ concentrations over East China in summer 2008 as observed from space and simulated with a chemistry transport model (CTM) based on a 2006 emission bottom-up inventory. The regional differences between observation and simulation can be mainly attributed to incorrect emission estimates, which for rising economies such as China are rapidly outdated by strong trends in emissions due to increasing economic activity [Richter et al., 2005; van der A et al., 2006, 2008].

Satellite measurements of atmospheric constituents can also be used for emission estimates, taking advantage of a homogeneous data set, which also provides information for regions where statistical data are sparse. Moreover, these inventories can be updated soon after the satellite data become available, therefore trends in emissions can be monitored on a regular basis. Emissions, however, cannot be measured directly, hence a CTM is needed to calculate a concentration field from a certain emission inventory. The difference between observed and modeled concentrations contains information for adjusting the underlying emissions. Due to transport away from the source, however, this inversion problem is computationally complex, because a single forward model run does not provide information on the (non-local) dependence of concentration on emissions.

In recent years, several studies have been published on estimates of NO$_x$ emissions from satellite observations, with a variety of techniques aimed at different spatial resolution and time scales, using different satellite instruments and CTMs.
If transport to neighboring grid cells can be neglected (either because of the large size of the grid cells or the short lifetime of the species) the problem is relatively easy and an approach as in Martin et al. [2003, 2006] can be applied, in which the NO$_2$ columns are linearly related to the NO$_x$ emissions. By optimal estimation, the new emissions are weighted between the top-down estimates and the a priori inventory [see also Jaeglé et al., 2005]. Zhao and Wang [2009] apply this method iteratively on a daily basis for a 70 km resolution, assuming that the impact of horizontal transport is accounted for indirectly by daily assimilation. Alternatively, Lin et al. [2010] make NO$_x$ emission estimates by looking at the concentration difference measured by two instruments at different times in the day. Knowing the lifetime of the species (which is taken from a CTM run), and the (prescribed) emission variation during the interval between the two observations, a new mean emission for a grid cell can be calculated.

At higher resolutions or for longer lifetimes, transport from the source becomes important and non-local sensitivities of concentration to emission must be calculated. When it is available, this can be done with an adjoint model code [e.g., Kurokawa et al., 2009; Stavrakou et al., 2008]. Another option is using the decoupled direct method [e.g., Napelenok et al., 2008] where sensitivities to emission areas are calculated by transporting them in the CTM through adapted transport equations. An alternative technique is used by Konovalov et al. [2006], who calculate sensitivities to neighboring grid cells by solving a set of linear equations obtained by model runs with random perturbations on the a priori emissions. All these methods have in common that they are time consuming, and concessions have to be made on the number of emission areas or the distance of transport which is taken into account.

Popular inversion methods are the Kalman filter (applied in, e.g., Napelenok et al. [2008]), in which an emission analysis is done at each observation time, and 4DVAR (applied in e.g., Kurokawa et al. [2009]), in which an emission analysis is done for a larger assimilation time window containing observations at different times. Both methods take the error covariances and their temporal evolution of the observations and the model into account. By using the ensemble Kalman filter (EnKF) technique [Evensen, 1994], used in an NO$_x$ emission estimation setting by Miyazaki et al. [2012], there is no need to calculate explicitly the sensitivities nor the evolution of the emission covariance. The probability density function of the latter is approximated by the spread of an ensemble (typically generated by 20–40 forward model runs) from which the emission estimate and its covariance can be calculated.

The Daily Emission estimates Constrained by Satellite Observations (DECSO) algorithm is complementary to the techniques above, aiming at emission resolutions where transport away from the source cannot be neglected (and fast inversion schemes as in Martin et al. [2003] cannot be applied), and at temporal resolutions requiring short calculation times (much faster than established techniques such as EnKF). More specifically, it was developed with the following requirements in mind:

1. The inversion can be applied to mesoscale emission inventories, and works at least for short-lived chemical species.
2. The inversion is fast enough to enable daily assimilation of satellite observations.
3. The inversion is able to detect new emission sources which are not included in the a priori emission inventory.
4. The inversion method converges sufficiently fast toward the new emission levels to enable short-term emission trend analysis.

The DECSO algorithm contains a new method of calculating local and non-local sensitivities of concentration to emission based on a single forward model run, without using adjoint model code or perturbation techniques (see Section 3). A Kalman filter is used to derive the emissions from the forward model run, the sensitivities and the covariances (Section 4). The inversion adds an emission update to the a priori emissions, instead of scaling them, enabling detection of unknown sources. Section 5 assesses the
performance of the algorithm with closed loop tests based on synthetic observations, while some first results based on real satellite data are presented in Section 6.

2. Observations, Model, and Emissions

[14] The DECSO algorithm will be applied to estimation of NO\textsubscript{x} emissions in East China, where due to the economic growth the emission rates are changing rapidly. We use satellite retrievals of tropospheric NO\textsubscript{2} columns by either the OMI or the GOME-2 instrument, and compare these by concentration simulations of the CHIMERE model on a 0.25 × 0.25 degree resolution, based on the INTEX-B emission inventory.

2.1. Satellite Data

[15] OMI and GOME-2 are both nadir viewing, spaceborne spectrometers which measure the solar radiation backscattered by the Earth’s atmosphere. GOME-2 is carried on the MetOp-A satellite which was launched in October 2006. It uses four channels to cover a spectral range from 240 to 790 nm, from which columns of O\textsubscript{3}, SO\textsubscript{2}, NO\textsubscript{2} and HCHO can be retrieved. Cloud pressure and cloud fraction are derived from the reflectivities in the oxygen A-band at 760 nm with the FRESCO algorithm [Koelemeijer et al., 2001]. The instrument scans with a mirror mechanism 1920 km across track, therefore having near-global daily coverage. In forward scan its footprint is 80 × 40 km\textsuperscript{2}; overpass time is around 9:30 local time. OMI (on board the Aura satellite, launched on July 2004) measures in the spectral range from 270 to 500 nm with a spectral resolution of about 0.5 nm, from which columns of O\textsubscript{3}, SO\textsubscript{2}, NO\textsubscript{2} and HCHO can be retrieved. Cloud pressure and cloud fraction are derived from the O\textsubscript{2}-O\textsubscript{2} absorption feature at 477 nm [Acarreta et al., 2004]. The 114° viewing angle of the telescope corresponds to a 2600 km wide swath on the surface, enabling a daily global coverage of its measurements. Its spatial resolution is 24 × 13 km\textsuperscript{2} in nadir, and increases to 68 × 14 km\textsuperscript{2} at the swath edges (discarding the outer 4 pixels). Overpass time is around 13:30 local time.

[16] OMI has the advantage of a higher spatial resolution and daily global coverage, and GOME-2 has the advantage of a stronger anthropogenic signal of NO\textsubscript{2} due to its earlier overpass in the day [Boersma et al., 2008]. For both instruments, the tropospheric NO\textsubscript{2} columns are taken from the KNMI retrieval product (version 2), described by Boersma et al. [2011] and available through the TEMIS portal (http://www.temis.nl/).

[17] Not all satellite retrievals are included in the data assimilation. Retrievals at cloudy conditions (cloud fractions larger than 20%) are filtered out, to reduce the influence of the modeled NO\textsubscript{2} column below the clouds in the retrieval. For bright surfaces such as snow (surface albedo larger than 20%) the cloud fraction from the cloud algorithm becomes uncertain, hence these retrievals are also discarded. Retrievals with clouds below 800 hPa are filtered out because the possible intersection of the cloud with the NO\textsubscript{3} bulk makes the retrieval too sensitive for the exact cloud height. For OMI, the large outer 4 pixels at either side of the swath are filtered out, as are pixels affected by the row anomalies (which appear since June 2007). For GOME-2 we discard the backscan pixels (which are too large), and retrievals in the zoom-mode of the instrument (which are of unknown quality).

2.2. Chemical Transport Model and Emission Inventory

[18] The CHIMERE multiscale model [Schmidt et al., 2001; Bessagnet et al., 2004] is primarily designed to produce daily forecasts of ozone, aerosols and other pollutants and make long-term simulations for emission control scenarios. In the presented configuration, CHIMERE has been implemented over East Asia (18°N to 50°N and 102°E to 132°E), simulating the atmosphere in 8 layers up to 500 hPa, with a horizontal resolution of 0.25° × 0.25°. The meteorological data is taken from the deterministic forecast of the European Centre for Medium-Range Weather Forecasts (ECMWF), which is given on 91 atmospheric layers for a horizontal resolution of approximately 25 × 25 km\textsuperscript{2}. The boundary conditions for the domain border and top are taken from monthly climatologies (no nested run is performed). To reduce the effect of boundary values on the region of interest, the domain boundaries have been chosen over relatively unpopulated areas. In Europe, CHIMERE has been extensively intercompared to other urban air quality models [e.g., Vautard et al., 2007] and evaluated against ground-based measurements and satellite data [e.g., Blond et al., 2007]. For China, validation results can be found in Mijling et al. [2009].

[19] As a starting point for our estimates, we use the recent INTEX-B emission inventory by Zhang et al. [2009a, also Q. Zhang, personal communication, 2009]. It covers Asia on a 0.2° × 0.2° resolution, containing the yearly totals of SO\textsubscript{2}, NO\textsubscript{x}, CO, VOC, PM\textsubscript{10}, PM\textsubscript{2.5}, BC, and OC by four sectors (power, industry, residential, and transportation) for the year 2006. This inventory has been successfully validated during the INTEX-B campaign [Zhang et al., 2008]. Biogenic emissions are not included, as for the selected domain the anthropogenic signal is supposed to be dominant. The NO\textsubscript{x} emission totals are interpolated to the model grid and distributes over three chemical components: NO (90% of the NO equivalent mass of NO\textsubscript{3}), NO\textsubscript{2} (9.2%), and HONO (0.8%). The emissions are disaggregated to hourly values using sector-specific weekly and diurnal factors; no monthly cycle is postulated.

3. The DECSO Algorithm

[20] We consider a time interval \( t = [0, T] \) between two data assimilation moments, in our case the 24 h period between two overpasses of the satellite instrument. At the core of the algorithm is the CTM which calculates the concentration fields of NO and NO\textsubscript{2} at \( t = T \) from the initial fields at \( t = 0 \), given the meteorological conditions and a certain emission database. After the model run, however, there is no information available of the sensitivity of the final NO\textsubscript{x} column concentration on the emission field; information needed by the Kalman filter for the emission inversion. Below, we will derive a simplified 2D transport equation which describes this relation analytically. It uses trajectory analysis to transport NO\textsubscript{x} columns over the model domain (Section 3.2). The aging of the NO\textsubscript{x} column is described by an effective lifetime, chosen in such way that it minimizes the calculated concentrations and the simulated concentrations by the CTM.
at \( t = T \) (Section 3.3). The concentration fields from the CTM are also used to construct the forecasted tropospheric \( \text{NO}_x \) column concentration, by interpolation toward the satellite footprint, extension with a climatological column to the tropopause, and application of the averaging kernel from the satellite retrieval method (Section 3.4).

### 3.1. Derivation of the Simplified 2D Transport Equation

In order to find the sensitivity of the \( \text{NO}_x \) concentrations to emission changes, we will derive a simple description of the transport of \( \text{NO}_x \) columns over the model grid with use of trajectory analyses neglecting the height dimension. Because of the rapid cycling between \( \text{NO} \) and \( \text{NO}_2 \), and the shift toward \( \text{NO}_2 \) during nighttime due to lack of photolysis, our analysis is based on the bookkeeping of \( \text{NO}_x \) which has been emitted at a certain moment in the time interval and subsequently transported during the remaining time. Analogous to equation (4), and integrating over all time dependent emission contributions \( \delta_i(t) \) during \( T \) (see Figure 2), we derive the expression:

\[
\delta_i^\text{H}(T) = \sum_j \frac{a_j}{a_i} \Omega_j(T) \delta_j(0) e^{-k_j(T-t)dt}
\]

in which the transport kernel \( \Omega_j(T) \) represents the fraction of the area of cell \( j \) which is transported to cell \( i \) during time interval \( T \). To ensure mass conservation, the equation is scaled with the area \( a \) of the corresponding grid cells. For a regular grid over latitude and longitude \( a \) will depend on latitude \( \phi \) and \( a = \cos(\phi)/\cos(\phi_i) \). During transport, the number of \( \text{NO}_x \) molecules in the plume will decay exponentially because of aging. We assume that the \( \text{NO}_x \) lifetime in the plume can be described by an effective lifetime \( \tau_j \) from its source location. Including aging in equation (3) we find for the concentration (written in terms of the reciprocal lifetime \( k \)):

\[
\delta_i^\text{H}(T) = \sum_j \frac{a_j}{a_i} \Omega_j(T) \delta_j(0) e^{-k_j(T-t)dt}
\]

The inversion only adjusts time-averaged emissions \( \delta_i \); the diurnal and weekly cycle of the emissions are described by an emission modulation function \( f(t) \), which is copied from the chemistry transport model:

\[
\delta_i(t) = f(t)\delta_i, \quad \langle f(t) \rangle_{\text{1week}} = 1
\]

Substituted in equation (5) this results in:

\[
\delta_i^\text{H}(T) = \sum_j \frac{a_j}{a_i} \left( \int_0^T e^{-k_j(T-t)}dt \right) \delta_j
\]
and $k$ can be interpreted as the $j$ in terms of $H$ to decay by a factor of $\frac{1}{C_0}$ emission in cell $W$ of NO\textsubscript{x} at time $t_0$ becomes diagonal, and the simplified 2D transport equation simplifies to:

$$ c_i(T) = \exp(-k_it) c_i(0) + \int_0^T \exp(-k_it)e_j(T-t)dt $$

which is the solution of the differential equation of a box model, having a loss term proportional to the concentration, and a time dependent emission:

$$ \frac{dc}{dt} = -kc + e(t) $$

Equation (9) describes the sensitivity of the NO\textsubscript{x} concentration in cell $i$ on the NO\textsubscript{x} emission in cell $j$. However, we are interested in the sensitivity $H_{ij}$ of the NO\textsubscript{y} concentration on the NO\textsubscript{x} emission (which is used in Section 3.4):

$$ H_{ij}^\ast = \frac{\partial c_j^{NO_2}}{\partial c_i^{NO_x}} $$

From the CTM simulation we obtain the NO\textsubscript{2}/NO\textsubscript{x} ratio $\gamma_i$ for all cells $i$ at the moment of assimilation $T$:

$$ c_i^{NO_2}(T) = \gamma_i c_i^{NO_x}(T), \quad 0 \leq \gamma_i \leq 1 $$

From this and equation (9) we get the sensitivities for NO\textsubscript{2} column concentrations to NO\textsubscript{x} emissions, both at the model grid:

$$ H_{ij}^\ast(k_i) = \gamma_i H_{ij} = \gamma_i \frac{a_j}{a_i} \int_0^T e^{-k_it} \Omega_{ij}(t)f_j(T-t)dt $$

### 3.2. The Transport Kernel

Driving force behind the transport of NO\textsubscript{x} is the time and space dependent wind field. In the simplified 2D transport equation the transport is described by matrix function $\Omega(t)$, whose column $j$ can be interpreted as the advection of the tropospheric column of NO\textsubscript{x} of grid cell $j$ over the model grid during the time interval $[T-t, T]$. We discretize $\Omega(t)$ at $n_T$ instants $t_n$, equally distributed over the assimilation time interval:

$$ \Delta t = T/n_T, \quad t_n = \left( n - \frac{1}{2} \right) \Delta t, \quad n = 1, \ldots, n_T $$

If $n_T$ is sufficiently large such that the transport distance is small during $\Delta t$ (at wind speed $v$ and grid cell size $\Delta x$ this implies $\Delta t \ll \Delta x/v$ must hold), the integral in equation (9) can be approximated by a summation over sparse matrix operations:

$$ H_{ij}(k_i) \approx \frac{a_j}{a_i} \sum_{t_n} e^{-k_it} \Omega_{ij}(t_n)f_j(T-t_n)\Delta t $$

The matrix elements $\Omega_{ij}$ are calculated for each $t_n$ by trajectory analysis, which must take the divergence of the wind field and the height dependence of wind and NO\textsubscript{x} concentration into account. In remote areas, for example, the bulk of the NO\textsubscript{x} tends to be located higher up in the troposphere (mainly because of uplifting and washing out of the NO\textsubscript{x} in the lower troposphere) where wind flows generally faster. A practical method for fast calculation of $\Omega_{ij}$ is by dividing the source grid cell $j$ in a large amount of $n_S$ randomly permuted sub grid cells. For each sub grid cell a trajectory is calculated, in order to sample how the content of the grid cell is distributed over the model grid. Each of the $n_S$ trajectories has a unique injection height, distributed according to the particle density in the NO\textsubscript{x} column, after which it is transported along the time dependent, two-dimensional wind field (based on the same wind fields that are used in the CTM run) vertically interpolated to this height. Then $\Omega_{ij}(t)$ can be found by counting the number of sub cells which originate in cell $j$ at time $T-t$ and are within cell $i$ at time $T$, see Figure 3.

As a trade-off between accuracy and calculation time, results in this paper are obtained by choosing $n_S = 10 \times 15$ sub cells and $\Delta t = 0.5$ h.

### 3.3. The Effective Lifetime

Equation (9) gives an expression for the sensitivity which depends on the lifetime $\tau$ of NO\textsubscript{x}, the time for emitted NO\textsubscript{x} to decay by a factor of $e$. The variation in lifetime, however, is too large to simply assume a fixed value for $\tau$. The lifetime of NO\textsubscript{x} can vary between several hours and several days, depending on factors such as temperature, sunlight, precipitation, altitude, and presence of other chemical species and aerosols. This dependence is too intricate to make a straightforward parameterization of $\tau$ which could be used in look-up tables. Instead, we use the results of the forward chemical transport model run to retrieve information on the NO\textsubscript{x} lifetime.

The matrix elements of $G$ and $H$ in equations (8) and (9) depend on the unknown reciprocal lifetimes $k$ associated with the grid cells. The residue $\tau$ is defined as the difference between the column concentrations $c^{\ast\text{CTM}}$ calculated with the
CTM and the concentrations calculated with the simplified 2D transport equation, at $t = T$:

$$r(k) = c_{\text{CTM}}(T) - (G(k)c_{\text{CTM}}(0) + H(k)e)$$  \hspace{1cm} (17)$$

The effective lifetime field is found by minimizing the residue for $k$ numerically, as explained in Appendix A. This is an ill posed problem: the exponential dependence in $k$ of the matrix elements and the reciprocal dependence of $\tau$ of $k$ make the residue function relatively insensitive for variations in both low and high values of $\tau$. Therefore we apply a Tikhonov regularization, demanding smoothness of the solution field as an a priori constraint. Unrealistic high or low values are corrected by constraining the lifetime $\tau$ between 2 h and 48 h. The lifetimes are calculated numerically, using the lifetime results from the previous day as an initial field.

Figure 4 shows the averaged fit results for the summer and winter period in 2008. The observed spatial variability results from the dependence of NO$_x$ lifetime on factors as latitude (photolysis rate and temperature), pollution composition, and meteorology. Especially in summer, the effective lifetimes are found lower for the OMI window than the GOME-2 window, which can be understand with equation (10). With $kT$ large and $e(t)$ large (in source areas), the second term is dominant, so the lifetime fit depends...
mainly on $e(t)$ and $c(T)$, the latter being significant lower at OMI overpass time due to the increased OH-producing photolysis rate in the early afternoon. In wintertime, the lifetime of the aged NO$_x$ plume in remote areas can be longer than 24 h, caused by low OH background concentrations and the bulk residing in higher layers of the troposphere, where it is less sensitive to uptake by cloud droplets and wet deposition. In populated areas the NO$_x$ bulk is closer to the surface, and the effective lifetime is found lower. For the OMI time window, we find a lifetime range of 3.3–12.8 h for Beijing (39.9°N), 4.3–7.7 h for Shanghai (31.2°N), and 6.4–7.0 h for Hong Kong (22.3°N); the smallest values found in summer, the highest in winter. Although the results for Hong Kong are higher than found by Beirle et al. [2011] for the Pearl River Delta (3–4 h), they agree in their small seasonal variation. Apart from weaker fluctuations at high latitudes, our results correspond well with the CTM calculations of Martin et al. [2003]. They find 5–6 h at 20°N, 5–8 h at 30°N, and 3–15 h at 40°N.

### 3.4. Comparing Model Simulations With Satellite Observations

For the data assimilation we will compare simulated NO$_2$ columns with observed NO$_2$ columns. The CHIMERE model calculates daily tropospheric columns up to 500 hPa. In order to extend the NO$_2$ profiles from the model ceiling to the tropopause, we add a climatological partial column for this part of the free troposphere. This climatology was compiled from a 2003–2008 run of the global chemistry transport model TM5, described in Huijnen et al. [2010], at a $2° \times 3°$ resolution and 34 atmospheric layers. Although in populated areas the added free tropospheric column contributes only a few percent to the total tropospheric column, it can account for up to 50% in remote areas, where the tropospheric NO$_2$ column is small.

The number of observations $m$ per day is smaller than the number of grid cells $n$, because observations are filtered, and the size of the satellite footprint is generally larger than a grid cell. For our domain $n = 15,609$, and contains typically 500–900 filtered GOME-2 retrievals; for OMI retrievals $m$ is in the range of 4500–7500. We interpolate the simulated grid concentrations to representative values for the satellite footprint, after which the averaging kernel from the retrieval method (important for a direct comparison between simulation and observation [see Eskes and Boersma, 2003]) can be applied directly.

We construct a representative vertical NO$_2$ profile $p_i$ for the footprint area by taking a weighted average of the modeled vertical profiles $p_j$ in contributing grid cells, taking weight factor $W_j$ from the surface fraction of grid cell $j$ which is covering footprint $i$:

$$p_i = \sum_j W_j p_j^* = W p^*, \text{with } W_j \text{ normalized } \sum_j W_j = 1$$  \hspace{1cm} (18)

Afterwards, the interpolated profile $p_i$ is rebinned to the model layers used in the tropospheric NO$_2$ retrieval algorithm, so the averaging kernel $A_i$, can be applied to get the modeled, corrected tropospheric NO$_2$ column at satellite footprint $i$:

$$y_i = A_i p_i$$  \hspace{1cm} (19)

This is the observable quantity. An example of how these simulated tropospheric columns relate to the observed columns can be seen in Figure 5.

For the inversion we will also need the Jacobian of the model, i.e., the matrix $H$ which represents the linearization of the model around a certain emission field $e$:

$$\Delta y = H \Delta e$$  \hspace{1cm} (20)

$H$ describes the sensitivity of the modeled observations $y$ (in observation space) to changes in model emissions $e$ (in state-space). We want to find an expression of $H$ in terms of the model sensitivities we found in equation (14). The gridded concentrations are interpolated to observation footprints by applying the interpolation matrix $W$ from equation (18):

$$\Delta \hat{y} = W H^* \Delta e$$  \hspace{1cm} (21)

in which $\Delta \hat{y}$ represents the change of the total tropospheric NO$_2$ columns in observation space, simulated by our model. Note that there is no information on the vertical profiles, which complicates a direct application of the averaging kernels. Therefore we assume that an emission change $\Delta e$ does change the total column value $\hat{y}_i$ for modeled observation $i$, but does not change its vertical profile shape, so we can write the change in averaged, modeled observation in terms of the change of modeled observations:

$$\Delta \hat{y}_i = \alpha_i \Delta \hat{y}_i, \alpha_i = \frac{\Delta p_i}{\hat{y}_i}$$  \hspace{1cm} (22)
[50] Combining equations (21) and (22), we get the expression for the Jacobian, in terms of the model sensitivities, the interpolation matrix, and the averaging kernels:

$$ \Delta y_i = \alpha_i \sum_j W_{ij} H_{jk} \Delta e_k \Rightarrow H = \text{diag}(\alpha) W H^* $$  \hspace{1cm} (23)

in which $H^*$ is the $(n \times n)$ model sensitivity matrix described by equation (14) for NO$_2$ columns to NO$_x$ emissions on the model grid, and $H$ is the $(m \times n)$ sensitivity matrix for NO$_2$ column observations to gridded NO$_x$ emissions, to be used in the Kalman equations below.

4. The Kalman Filter for Emission Estimation

[51] The relation between column concentration difference vector $\Delta y$ and emission update vector $\Delta e$ can be written as $\Delta y = H \Delta e$. Note that solving $\Delta e$ from $\Delta y$ and $H$ is an underdetermined problem. Furthermore, the errors in $y$ are large, and they would propagate non-locally in the solution, causing strong fluctuations in the emissions if this assimilation scheme is applied iteratively. To deal with these issues we use the Kalman filter, which calculates for every assimilation step the most probable emission field and its covariance, taking into account the errors in the modeled emissions and representation, and the errors in the observed concentrations. Due to the nonlinearity of the problem the extended Kalman filter is used, which linearizes about the current mean and covariance:

State vector forecast $\tilde{e}^i(t_{i+1}) = M_t [e^i(t_i)]$  \hspace{1cm} (K1)

Error cov. forecast $P^i(t_{i+1}) = M_t P^i(t_i) M_t^T + Q(t_i)$  \hspace{1cm} (K2)

Gain matrix $K_t = P^i(t_i) H_t^T [H_t P^i(t_i) H_t^T + R_t]^{-1}$  \hspace{1cm} (KG)

State vector analysis $e^i(t_i) = e^i(t_i) + K_t (y_i - H_t [e^i(t_i)])$  \hspace{1cm} (K3)

Error cov. analysis $P^i(t_i) = (I - K_t H_t) P^i(t_i)$  \hspace{1cm} (K4)

The interpretation of the quantities is as follows:

- $e^i$, $e^f$ analysis and forecast of the NO$_x$ emissions;
- $P^i$, $P^f$ error covariance matrices ($n \times n$) of the forecasted emissions $e^f$ and the analysis of the emissions $e^a$;
- $M_t$ model describing temporal evolution of the emissions. $M_t$ applied on the true state is assumed to introduce no bias: $e^f(t_{i+1}) = M_t [e^a(t_i)] + e_i$, in which the noise is normally distributed around 0 with covariance $Q$;
- $Q$ covariance matrix ($n \times n$) of the modeled emissions (see Section 4.2);
- $M$ ($n \times n$) matrix representation of emission model $M_t$;
- $y$ observations of tropospheric NO$_2$ columns;

$H$ observation operator which relates the emissions of NO$_x$ in the model grid to the observable tropospheric column concentrations of NO$_2$. $H$ applied to the true state is assumed to introduce no bias: $y_i = H_t [e^a(t_i)] + e_i$, in which the noise is normally distributed around 0 with covariance $R$;

$R$ ($m \times m$) matrix describing the covariances of the observation operator $H$ (see Section 4.1);

$H$ Jacobian of the observation model $H_t$, linearized around state $e = e^f$; $y = y^f + H(e^f - e^a)$, describing how the tropospheric NO$_2$ column of observation $i$ changes when the NO$_x$ emission $e$ in grid cell $j$ is changed.

Note that we use the Kalman equations to update the trace gas emission from the observation of its concentration. Therefore the attribution of the quantities in the Kalman formalism differs from e.g., data assimilation of ozone concentration fields (e.g., by Eskes et al. [2003]). In the latter case the state vector represents the modeled trace gas concentrations, $M$ describes the evolution of this field by the chemical transport model, $y$ represents the trace gas retrievals, and $H$ the representation of these retrievals on the model grid.

4.1. Covariance of the Observation Minus Forecast

[52] The covariance of the observation minus forecast of the column concentration consists of three independent components:

$$ \Sigma_{\text{OmF}} = \Sigma_{\text{obs}} + \Sigma_{\text{repr}} + HP^f H^T = R + HP^f H^T $$  \hspace{1cm} (24)

[54] The observation error of the tropospheric NO$_2$ column is composed of errors by the measurement noise of the satellite instrument and the spectral fitting, errors related to the separation of the troposphere and the stratosphere, and errors due to retrieval method parameters, such as clouds, surface albedo and a priori profile shape [Boersma et al., 2004]. The second part describes representation error originating from an inaccurate CTM (due to errors in e.g., meteorology or chemistry scheme), and errors introduced by adding a climatological free tropospheric column, and interpolating grid values to the satellite footprint. Together with the observation errors they are contained in covariance matrix $R$. The last part of the OmF error describes how errors in the emission estimation propagate into the simulated column concentrations. The sensitivity matrix $H$ is assumed to be exactly known; the error made by the approximating $H$ with the simplified 2D transport equation is added to $R$.

[55] It is the balance between $R$ and $HP^f H^T$ which determines how much information from the observed concentration difference is used to update the emission estimates. We work out a practical method to estimate the covariance matrix $R$. By neglecting spatial correlations the OmF error $\sigma_{\text{OmF},i}$ for observation $i$ can be written, analogous to equation (24) as

$$ \sigma_{\text{OmF},i}^2 = \sigma_{\text{obs},i}^2 + \sigma_{\text{repr},i}^2 + \sigma_{\text{prem},i}^2 \equiv \sigma_{R,i}^2 + \sigma_{\text{prem},i}^2 $$  \hspace{1cm} (25)

in which $\sigma_{\text{prem},i}^2$ is the propagated emission variance of matrix $HP^f H^T$. The observation error $\sigma_{\text{obs},i}$ is known from the
satellite product: typically they have a dominating absolute error at low values (around \(0.5 \times 10^{15} \text{ molecules/cm}^3\)), and a dominating relative error at high values (around \(30-45\%\)). The representation error \(\sigma_{\text{repr}}\) is unknown, but is assumed to be relative to the simulated tropospheric column concentration \(y\) with a fixed \(\varepsilon_{\text{rel}}\):

\[
\sigma_{\text{repr},i} = \varepsilon_{\text{rel}} y_i
\]  

[56] For each assimilation a large quantity of observations \(y_{\text{obs}}\) and forecasts \(y\) are available, enabling the calculation of \(\sigma_{\text{repr}}\) from OmF statistics, using the reduced \(\chi^2\) criterion:

\[
\chi^2_{\text{rel}} = \frac{1}{m} \sum_{i} \left( \frac{y_{\text{obs},i} - y_i}{\sigma_{\text{OmF},i}} \right)^2
\]  

[27] We calculate \(\varepsilon_{\text{rel}}\) such that \(\chi^2_{\text{rel}} = 1\), meaning that the variation in OmF is well described by its error \(\sigma_{\text{OmF}}\): the distribution of \((y_{\text{obs}} - y)/\sigma_{\text{OmF}}\) will be Gaussian around 0 with standard deviation 1. By substituting (25) and (26) in (27) we find the equation:

\[
f(\varepsilon_{\text{rel}}) = \sum_{i=1}^{m} \left( \frac{y_{\text{obs},i} - y_i}{\sigma_{\text{OmF},i}} \right)^2 + \sigma_{\text{obs},i}^2 + \sigma_{\text{prec},i}^2 - m = 0
\]  

[58] We can find the root of \(f\) numerically using its derivative to \(\varepsilon_{\text{rel}}\) in Newton’s method, resulting in \(\varepsilon_{\text{rel}} = 33\%\) for OMI observations in May 2008, see Figure 6.

[59] With the daily estimated value of \(\varepsilon_{\text{rel}}\) we calculate \(\sigma_{\text{repr},i}\) according to (25). From these errors we construct the covariances analogous to Eskes et al. [2003], by decomposing the error covariance matrix as

\[
R = \text{diag}(\sigma_R) \text{diag}(\sigma_R)
\]  

Figure 6. (left) Reduced chi-square test for OmF and its error for OMI observations of May 2008 as function of the relative representation error. (right) Distribution of OmF/\(\sigma_{\text{OmF}}\), with \(\sigma_{\text{repr}} = 33\%\). The dotted line represents the Gaussian distribution around 0 with standard deviation 1.

\[
\text{in which } C \text{ is a correlation matrix, which elements only depend on the distance between two observation footprints: } C_{ij} = g(r_{ij}). \text{ Theoretically, function } g \text{ can be estimated from the statistics on how the correlation between observations and forecasts decrease over distance, but here we model } g \text{ exponentially dependent on distance:}
\]

\[
g(r_{ij}) = \exp(-r_{ij}/L)
\]  

[30] in which \(L\) is the correlation length. Best inversion results are obtained by taking \(L\) small with respect to the footprint size (we will use \(L = 10\) km); in this case the improved condition of the matrix which is inverted and the Gaussian approximative solution (29) suppresses spatially oscillating solutions. Correlations between footprints at a distance larger than \(6L\) are considered insignificantly: corresponding matrix elements \(C_{ij}\) are set to zero, resulting in a sparse covariance matrix \(R\).

4.2. Emission Covariances and Inversion Behavior

[60] For the Kalman filter approach we need to assess the emission model \(M\), its error covariance \(Q\), and the initial covariance of the emission database \(P(t = 0)\). For these matrices it is important to find realistic estimates, which will optimize the assimilation for both convergence speed and noise reduction.

[61] Anthropogenic emission trends, if present, are usually very gradual, justifying a persistent emission model, stating that tomorrow’s emissions are equal to today’s emissions. This implies that the emission model reduces to the identity matrix: \(M = I\). Note that our algorithm adjusts weekly averaged emission totals; the diurnal and weekly cycle is modulated already by the CTM, see equation (6).

[62] After multiple iterations the error covariance analysis \(P(t)\) becomes independent of its initial value \(P(t = 0)\). However, to get optimal assimilation results directly from the start, we model \(P(t = 0)\) from the error distribution of the
Figure 7. Emission inversion behavior for different emission model errors (rows) and for different emission scenarios (columns). The dotted lines indicate the true emission values.
emission analysis after a test run of 60 days, sufficiently long for its errors to be close to their expected limit values.

[63] To assess the influence of the emission covariance matrix \( \mathbf{Q} \) on the inversion behavior we analyze the Kalman equations for the simple case where the concentration \( y \) only depends on one emission source \( x \) with a constant sensitivity factor \( h \). The Kalman filter reduces to the following scalar equations, and matrix \( \mathbf{Q} \) becomes a scalar quantity \( q \), dictating how much the error of the emission increases between two assimilation moments. We find an expression for the evolution of the error analysis \( \sigma^2_i \) in terms of its predecessor \( \sigma^2_{i-1} \):

\[
\sigma^2_i = \frac{\sigma^2_{R,i}}{h^2(\sigma^2_{i-1} + q^2) + \sigma^2_{R,i}} \left( \sigma^2_{i-1} + q^2 \right)
\]

[64] We can see with this formula that the limit value of the assimilation error \( \sigma^2 \) depends on the sensitivity \( h \) (a higher \( h \) results in faster convergence to a lower \( \sigma^2 \), the observation and representation error \( \sigma_R \) (a lower \( \sigma_R \) results in a lower \( \sigma^2 \)), and emission error increment \( q \). \( q \) is the only unknown parameter and its value will affect the noise and convergence in the assimilation. A high emission uncertainty \( q \) results in a large noise on the assimilated emission, and a low uncertainty \( q \) leads to slow convergence.

[65] We model \( q \) with a dominating relative error \( \epsilon_{\text{rel}} \) at low emissions, shifting to a dominating relative error \( \epsilon_{\text{rel}} \) at high emissions:

\[
q(x) = \epsilon_{\text{abs}} \exp(-\epsilon_{\text{rel}} x / \epsilon_{\text{abs}}) + \epsilon_{\text{rel}} x
\]

[66] Note that the absolute error component is essential to be able to pick up changing emissions in areas where zero emission are defined.

[67] We have tested different settings for \( q \) for different emission scenarios, see Figure 7. For all experiments we used observations with a realistic error model for \( \sigma_R \), containing an absolute and relative error analogous to equation (32), with \( \epsilon_{\text{abs}} = 0.5 \times 10^{15} \) molecules/cm\(^2\) and \( \epsilon_{\text{rel}} = 70\% \). For the sensitivity we used \( h = 0.25 \). For the initial analysis error \( \sigma_0^2 \) we take the expected limit value as calculated by (31). In the first series of experiments we model \( q \) with \( \epsilon_{\text{abs}} = 0.1 \times 10^{15} \) molecules/cm\(^2\)/h and \( \epsilon_{\text{rel}} = 10\% \). Experiments for strong emission sources in Figures 7a, 7b and 7c show that the large relative error is propagated to a large emission noise. This is reduced in the second series of experiments by taking \( \epsilon_{\text{abs}} = 0.1 \) and \( \epsilon_{\text{rel}} = 5\% \), see Figures 7f, 7g, and 7h. The low signal-to-noise values in Figures 7i and 7j is caused by the large absolute error. This is improved in a third series of experiments with \( \epsilon_{\text{abs}} = 0.02 \) and \( \epsilon_{\text{rel}} = 5\% \), see Figures 7n and 7o, although at the cost of a slower convergence of emerging strong emission sources in Figures 7m. The occurrence of emission noise in an emission-free area (Figure 7o) (which is positively biased because we do not allow negative emissions) can therefore be controlled by lowering \( \epsilon_{\text{abs}} \) although it should not be taken too small, as in the last series where \( \epsilon_{\text{abs}} = 0.01 \), where convergence speed of new strong (Figure 7r) and weak (Figure 7s) sources become unacceptably slow.

[68] For our assimilation setup we construct matrix \( \mathbf{Q} \) analogous to (29) and (30), modeling its errors according to (32) with \( \epsilon_{\text{abs}} = 0.02 \times 10^{15} \) molecules/cm\(^2\)/h and \( \epsilon_{\text{rel}} = 5\% \), and allowing for weak covariances with its nearest neighbors by taking the covariance length \( L = 10 \) km.

### 4.3. The Kalman Gain

[69] After establishing the covariance matrices as described in the above sections, the Kalman equations can be solved numerically. Calculation time and storage space is reduced by making optimal use of the sparseness of the matrices. Point of concern is the inversion of the symmetric matrix \( \mathbf{A} = \mathbf{H} \mathbf{P}^T \mathbf{H}^T + \mathbf{R} \) in the Kalman Gain (equation (KG)), which generally is ill-conditioned and contains a null-space. Note that \( \mathbf{A} \) being symmetric and positive semi-definite implies that \( \mathbf{A} \) has real, positive eigenvalues. The eigenvalues spectrum of matrix \( \mathbf{A} \) consists of only a few large eigenvalues and many smaller ones, which makes solving the inverse of \( \mathbf{A} \) very sensitive to noise. This sensitivity issue can be avoided by approximating \( \mathbf{A} \) with a decomposition

\[
\mathbf{A} \approx \mathbf{U} \mathbf{U}^T \Rightarrow \mathbf{A}^{-1} \approx \mathbf{U}^{-1} \mathbf{U}^T
\]

in which \( \mathbf{A} \) is a diagonal matrix consisting of the \( m \) largest eigenvalues of \( \mathbf{A} \). Here we use the numerical software library ARPACK [Lehoucq et al., 1998] for a fast calculation of the largest eigenvalues and corresponding eigenvectors. Analogous to Segers et al. [2005], we take \( m \) such that the sum of the largest eigenvalues account for 98% of the value of the trace of \( \mathbf{A} \), with a practical maximum of 1800 eigenvalues. As a result the condition number (the ratio of the largest eigenvector to the smallest eigenvector) is reduced to an order of 1000.

[70] For large number of observations, solving the Kalman equations becomes dominant in the total calculation time. The time spent in the ARPACK algorithm to calculate \( m \), eigenvalues scales with \( m^2 \). The eigenvector matrix \( \mathbf{U} \) in equation (33) has dimension \( m \times m \), and is generally dense, which means that \( \mathbf{A}^{-1} \) is calculated with \( m \times m \) floating point operations. Another time consuming calculation is the Kalman Gain matrix from sparse \( m \times n \) matrix \( \mathbf{P} \mathbf{H} \mathbf{U}^T \) with dense \( m \times m \) matrix \( \mathbf{A}^{-1} \), which takes \( 2mnq \) floating point operations, where \( q \) indicates the mean number of nonzero values in a row of \( \mathbf{P} \mathbf{H} \mathbf{U}^T (q \ll m) \).

### 4.4. Calculation of the Analysis Covariance

[71] Correlations in the emission analysis are introduced because different emission grid cells contribute to an observation in the satellite footprint. Although the covariances can be strongly non-local (e.g., when fast winds transport long living NO\(_X\) over remote areas), normally the covarying emission area is localized within a certain distance of the grid cell. Therefore it is not necessary to calculate all \( n \times n \) elements of the analysis covariance matrix \( \mathbf{P} \mathbf{H} \mathbf{U}^T \). Instead we calculate the diagonal elements \( \mathbf{P}^n_{ii} = \sigma^2_{ii} \) using (K4). Off-diagonal, we only calculate the covariance for grid cells within a radius of 300 km. Only correlations larger than 0.01 are supposed to contribute significantly to the analysis:

\[
|\rho_{ij}| = \frac{P^n_{ij}}{\sigma_i \sigma_j} > 0.01
\]

[72] By neglecting all smaller correlations the analysis covariance matrix becomes a sparse matrix.
4.5. Emission Inventory Update

[73] After calculating new emissions we have to make certain assumptions to use this data to update the emission database. For the moment, we only adjust the NO\textsubscript{x} emissions, neglecting the possibility that a change in anthropogenic NO\textsubscript{x} is related to a change in other anthropogenic emissions such as particulate matter and CO. Furthermore, we assume that the emission change is due to anthropogenic sources. Since we do not know how the new NO\textsubscript{x} emissions can be attributed to the different sectors (power, industry, residential, and transportation in the used INTEX-B inventory), we scale emissions in all sectors by ratio, assuming that the introduced error (through different sectorial diurnal and weekly emission cycles) is small. Because in our method emissions are updated by addition instead of scaling, emissions can be attributed to grid cells without a priori emission. For new emission sources, where no sector information is known, the new NO\textsubscript{x} emissions are distributed evenly over all sectors. Finally, information about the injection height of the new emission cannot be inferred with the DECSO algorithm. Instead, emissions at all heights will be scaled by ratio.

5. Closed Loop Tests

[74] A closed loop test is performed to assess the performance of the DECSO algorithm. To do so, we implement emission scenarios by adapting the INTEX-B emission inventory locally. From these new “true” emissions synthetic NO\textsubscript{2} observations are constructed. These synthetic observations are calculated by (1) running CHIMERE with the adapted inventory, (2) extending the column concentrations to the tropopause, (3) interpolating the results on the satellite footprints from selected real observations of GOME-2 and OMI, and (4) adding typical instrumental noise (analogous to equation (32), with $\epsilon_{\text{abs}} = 0.5 \times 10^{15}$ molecules/cm\textsuperscript{2} and $\epsilon_{\text{rel}} = 30\%$). The sensitivity is made independent of height by taking a flat averaging kernel. By studying the difference between the true emissions and the assimilated emissions (based on the synthetic observations), we assess the characteristics of the assimilation algorithm by checking the sensitivity, the convergence speed, and the noise.

5.1. Emissions Increase in Populated Areas

[75] To test if our algorithm can detect an emission growth in a grid cell while the emissions in surrounding grid cells remain constant, we assume a 30% increase in NO\textsubscript{x} emissions in the grid cell containing the city of Jinan (population 6 million (2006)), from $8.43 \times 10^{15}$ molecules/cm\textsuperscript{2}/h. Values in the 8 surrounding grid cells are considerable lower, ranging from 0.5 to $2.8 \times 10^{15}$ molecules/cm\textsuperscript{2}/h.

[76] Figure 8 (top) shows the results of the closed loop test for synthetic observations without added noise. The
emissions converge almost monotonically to their true value, faster for OMI than for GOME-2, due to its larger number of daily observations. The more frequent sampling also results in a smaller emission analysis error. When noise is added (Figure 8, bottom) convergence becomes slower. Due to the instrumental noise, emissions can be pushed away from its equilibrium for a longer period, before it returns to its true value. The emission estimate for OMI footprints takes 25 days to converge to its true value, while for GOME-2 observations it takes more than 70 days.

To investigate whether the algorithm is able to detect an emission increase of an expanding mega city, we assume a scenario where NO\textsubscript{x} emissions of 8 grid cells in the Pearl River Delta increase by 30%, while the NO\textsubscript{x} emissions in the neighboring Hong Kong area (2 x 2 grid cells) decrease by 30%. Figure 9 shows the result of the assimilation after a 90 day period with synthetic OMI observations, where the assimilated emission inventory has been averaged over the last 30 days. Starting with the unchanged emission inventory, the algorithm is capable of reproducing the emission scenario.

5.2. Detection of a New Emission Hot Spot

To emulate the appearance of a new emission source in remote, clean areas (e.g., by the introduction of a new power plant or industry, or the occurrence of a wildfire), we
increase the NO\textsubscript{x} emission of a single grid cell in the Xilin Gol area of Inner Mongolia from 0.18 to 2.0 \times 10^{15} \text{molecules/cm}^2/\text{h}. Values in the surrounding grid cells remain unchanged, ranging from 0.03 to 0.18 \times 10^{15} \text{molecules/cm}^2/\text{h}. As can be seen in Figure 10 (top), the new emission is picked up readily from synthetic OMI observations. The emission analysis based on GOME-2 observations show significant slower convergence. This can be attributed to a larger footprint size which smears out the emission attribution: Figure 10 (bottom) shows that if the emission analysis of surrounding grid cells is taken into account, the area total converges within 40 days acceptably close to its new level.

[79] Note that, for unknown new sources in remote areas, the model will typically simulate the height distribution of an aged NO\textsubscript{2} plume: a very low NO\textsubscript{2} column concentration with a maximum higher up in the troposphere. The unrealistic vertical profile will overestimate the correction factor $\alpha$ in equation (22), leading to an underestimated emission update $\Delta e$ based on the OmF in equation (23) in the first run. However, the iterative application of the algorithm compensates for this effect: the new surface source (even if it is small) will shift the maximum of the modeled NO\textsubscript{2} profile rapidly down to the surface, resulting in more realistic emission updates in subsequent iterations.

6. First Results

6.1. NO\textsubscript{x} Emission Estimates for East China

[80] To assess the algorithm performance in the East China domain with real satellite data, we compare a model run based on a constant INTEX-B inventory (a-priori) with
a model run with estimated emissions based on OMI data from May to December 2008. The data assimilation clearly results in a better reproduction of the spatial features of the tropospheric NO$_2$ column concentration, see Figure 11. When the mean emission estimates are compared with the a priori inventory, several features can be noted: (1) increased emissions along the Yangtze River (inland from Shanghai) and the Pearl River (inland from Hong Kong); (2) decreased emission in North Korea; (3) detection of new emission hot spots in Inner Mongolia, due to the operation of new coal fired power plants [Zhang et al., 2009b]; (4) coastal ship tracks (not accounted for in the a priori inventory) between the Pearl River delta and the Taiwan Strait, and in the yellow sea, connecting Shanghai with the tip of Shandong Peninsula. The locations correspond with the AMVER-ICOADS shipping database [Wang et al., 2008]. Shipping routes in the Bohai Sea are diverse, hence they show up as an area source.

The Kalman filter keeps track of the error estimates of each grid cell emission. The error of the emission analysis depends on the sample frequency of the downwind concentration plume, the average sensitivity of the emission to the observed concentration, and the error growth of the emissions between assimilation moments (Section 4.2). In Figure 12 the emission analysis errors of the entire domain are plotted against the emissions, and are modeled according to equation (32). For OMI data assimilation, we find a dominating absolute error of $0.38 \times 10^{15}$ molecules/cm$^2$/h at low emissions, and a dominating relative error of 47% at high emissions. Data assimilation with GOME-2 results in larger emission errors due to the smaller amount of observations: $0.48 \times 10^{15}$ molecules/cm$^2$/h at low emissions, and 74% at high emissions. Note that when the emissions converge to their true state the analysis errors become over-estimated due to an overestimation of the emission model covariance $Q$. However, inflating the emission forecast covariance by a significant $Q$ at every assimilation cycle is necessary to avoid becoming insensitive to emission changes.

[81] The DECSO algorithm applied for East China with OMI observations from May—December 2008. (top) Comparison of the mean OMI observations of tropospheric NO$_2$ with colocated CHIMERE forecasts with and without assimilated emissions. (bottom) The a priori (INTEX-B) emission inventory, the mean assimilated emissions for this period, and the emission update, respectively.
individual Beijing grid cells in the OMI time series of Figure 13, the mean relative emission error is 58%. By taking an average over 4 grid cells, the mean error drops to 23%. By taking the negative covariances between the grid cells into account the emission analysis error drops further to 20%. For the GOME-2 time series, the errors are 91%, 47%, and 38%, respectively. The effect of the negative covariances is here stronger for GOME-2 due to the larger footprint size (see Section 5.2).

To study the effect of transport on the emission estimates we perform an additional run with the DECSO algorithm in which the transport to neighboring grid cells has been disabled by taking the unity matrix for the transport kernel in equations (8) and (9). This couples the column concentration locally to the underlying emissions. Note that this is basically similar to the method used by Martin et al. [2003], however, it differs on three important points: (1) it is applied iteratively, (2) its inversion uses a Kalman filter based on a forecasted emission error and the estimated OmF error, and (3) emissions are updated by addition instead of scaling.

Evaluating of how the emission estimates improve by taking transport into account is complicated by the lack of knowledge of the true emission inventory. Instead, we look at the improvement of the forecast capability for tropospheric NO$_2$. We calculate for all locations in the model domain the root mean square (RMS) of the OmFs, for OMI observations during October–December 2008, and for forecasted emissions estimated with and without transport. The observations are taken from OMI for the period October to December 2008. The RMS is weighted by the retrieval error, which means that for perfect simulations the RMS

![Figure 12](image1.png)

**Figure 12.** Scatterplot for emission analysis against its error, for all emissions in the domain at 1 December 2008. Results for (left) OMI and (right) GOME-2.

![Figure 13](image2.png)

**Figure 13.** Time series from May–December 2008 for NO$_x$ emission estimates over Beijing based on OMI (blue) and GOME-2 (red) observations of tropospheric NO$_2$. The shaded areas indicate the emission analysis error.
would be 1. Higher values indicate larger errors in the tropospheric NO\textsubscript{x} simulation.

Especially emission estimates in areas downwind of strong emission sources are affected when transport is neglected. We select all areas where the ratio between the mean column concentration and the mean emission is larger than 24 h (i.e., the amount of NO\textsubscript{x} cannot be explained by local emissions only), which covers 26% of the model domain, for OMI observations in October to December 2008. By taking transport into account in the emission estimates, 65% of this area improves its RMS of the error weighted OmF. Strong improvement is made for instance in North Korea, where NO\textsubscript{x} concentrations by local emissions are low compared with the inflow of NO\textsubscript{x} from China transported by westerly winds.

6.2. NO\textsubscript{x} Emission Trends for the Beijing Area

As a case study for the performance of the DECSO algorithm with real satellite data, we construct a time series of NO\textsubscript{x} emission estimates of the Beijing area, consisting of the average of 4 grid cells which are enclosed by Beijing’s sixth ring road, which spans 130 km and surrounds the city center at 15–20 km distance. Both OMI and GOME-2 estimates show the same emission trend; see Figure 13. For the period May–September 2008 this is a downward trend, mainly because of the effectiveness of the air quality measures taken by the authorities before and during the 2008 Olympic and Paralympic Games, from 8 August to 17 September [Mijling et al., 2009].

Due to cloudy conditions at the start of the Olympic Games, all satellite observations of NO\textsubscript{x} around Beijing are filtered out. At August 15, for the first time since the start of the Games the Beijing area is fully sampled by OMI, and the emission are updated toward the new low levels. For GOME-2 this happens at 23 August. If we compare the emission levels of the last week of August with the pre-Olympic emission levels in May–June 2008 (with mean emission levels of 8.1 and 7.8 $10^{15}$ molecules/cm\textsuperscript{2}/h for OMI and GOME-2 respectively), we find an emission reduction of 45% (OMI) and 50% (GOME-2). If a linear relationship between surface emissions and column concentrations is assumed [Martin et al., 2006], these numbers are lower than the 60% concentration reduction found by Mijling et al. [2009] would suggest. This implies that, if a linear relation between emission and concentration is assumed, the concentration drop in Beijing cannot be explained by local emission reductions alone. The non-local contributions to NO\textsubscript{x} over Beijing (from upwind sources outside the city) must have dropped by more than 60%.

After the events, NO\textsubscript{x} emission for the Beijing area starts to rise again, in accordance with the increase of in NO\textsubscript{2} concentrations observed by Witte et al. [2009]. In December 2008, emissions are at 75% of their pre-Olympic levels.

6.3. Tropospheric NO\textsubscript{2} Columns Over East China

The ability of the CTM to reproduce realistic concentration fields in populated areas is especially important for air quality forecast purposes. As stated in Section 1, the quality of the CHIMERE output depends strongly on the correctness of the emission data. Figure 14 compares the East Asian time series of observed tropospheric NO\textsubscript{2} concentration with time series of simulated concentration, with and
without data assimilation, by looking at the RMS of the error weighted OmFs. For populated areas, the simulations based on the assimilated NO\textsubscript{2} emission inventory result in a better agreement (i.e., lower RMS values) between simulated and observed NO\textsubscript{2} concentration, as expected. Also NO\textsubscript{2} concentrations for the Chinese coastal waters improve by attributing NO\textsubscript{2} emissions to this area (note that the INTEX-B inventory does not include shipping emissions).

7. Conclusion

[90] We developed a new algorithm, DECSO, which is designed for fast daily emission estimates from assimilated satellite observations of short-lived species. The algorithm was applied to the estimation of NO\textsubscript{2} emissions over East China on a 0.25 degree resolution from daily OMI and GOME-2 observations of tropospheric NO\textsubscript{2} column concentrations. Non-local relations between emissions and concentrations due to transport away from the source have been taken into account, and were obtained from a simplified 2D transport matrix equation. The errors introduced by the assumptions made in this equation are largely compensated for by fitting an effective lifetime which gives the best agreement between the concentration forecast of the transport equation and the forecast of the full CTM run. The CTM forecast is also used to produce simulated tropospheric NO\textsubscript{2} columns which are colocated with the observations. The difference between the observation and the simulation and the relation between emission and concentration are used in a Kalman filter to calculate the new NO\textsubscript{2} emission inventory.

[91] Closed loop tests with synthetic satellite observations show that the algorithm is capable of reconstructing new NO\textsubscript{2} emission scenarios from tropospheric NO\textsubscript{2} column concentrations. It is also capable of detecting new emission sources such as power plants and ship tracks which are unaccounted for in the a priori emission inventory. Using OMI and GOME-2 data, the algorithm is able to detect emission trends on a monthly resolution, such as during the 2008 Beijing Olympic Games. Furthermore, the tropospheric NO\textsubscript{2} concentrations calculated with the new emission estimates show better agreement with the observed concentrations, both in space as in time, facilitating the use of the algorithm in operational air quality forecasting.

[92] The advantage over techniques relating the observed NO\textsubscript{2} columns linearly to the underlying NO\textsubscript{2} emissions is that the DECSO algorithm takes transport to other grid cells into account, enabling emission estimates at higher resolution and at longer lifetimes (e.g., NO\textsubscript{2} in wintertime). Experiments for OMI observations in October—December 2008 show that the results improve when transport is taken into account, especially above areas where concentrations by local emissions are low compared to the inflow of NO\textsubscript{2} from upwind source areas. Also, the DECSO algorithm updates emissions by adding contributions, rather than by scaling existing emissions, which enables the detection of new sources that are unaccounted for in the a priori emission inventory. The main advantage over techniques using 4DVAR or ensemble Kalman filtering (EnKF) is the calculation speed of the algorithm, which facilitates e.g., operational application in NO\textsubscript{2} concentration forecasting at mesoscale resolution. Only one forward CTM run is needed (taking 1 time unit for CHIMERE, about 18 min in our setup), followed by the calculation of the inversion (taking another 1.1—2.8 time units, depending on the number of observations for that day). Emission inversion with EnKF or 4DVAR techniques typically takes 20—40 time units. Calculation times can be further decreased by e.g., simplifying the calculation of the transport kernel (at the cost of accuracy) or parallelization of the computer code.

[93] A drawback of daily emission inversions on a high spatial resolution, such as in the presented method, is the higher sensitivity to retrieval noise when compared to other methods for larger temporal or spatial scales. The high grid resolution inhibits the use of spatial averaged observations, and the daily inversion inhibits the use of time averaged observations, both having smaller associated errors. Strong day-to-day fluctuations, however, are avoided by the Kalman filter which acts as a low-pass filter.

[94] Given the instrument characteristics (observational noise, sampling frequency, and footprint size) (taking 1 time unit for CHIMERE, about 18 min in our setup), followed by the calculation of the inversion (taking another 1.1—2.8 time units, depending on the number of observations for that day). Emission inversion with EnKF or 4DVAR techniques typically takes 20—40 time units. Calculation times can be further decreased by e.g., simplifying the calculation of the transport kernel (at the cost of accuracy) or parallelization of the computer code.

Appendix A

[97] To find a best fitting reciprocal lifetime field k, which minimizes the difference between the column concentrations calculated by the CTM and the column concentrations calculated by the simplified 2D transport equation (equation (17)),
we introduce a cost function $J$ which sums the squares of the error-weighted residues:

$$J(k) = r(k)^T r(k) \quad (A1)$$

Generally, the minimum of $J$ cannot be found analytically due to its nonlinear dependence on $k$, and has to be solved numerically. The dependence of the matrix elements of $G$ and $H$ on $k$ makes the evaluation of gradient $\nabla_k J$ expensive, which discards methods such as steepest descent or conjugate gradient. The dependence of the matrix elements on $k$ can be circumvented by a linearization around $k^0$, $k = k^0 + \Delta k$. We get for matrix $G$:

$$G_y(k^0 + \Delta k) = G_y(k^0) + \Delta G_y = \left[ \begin{array}{c} \frac{\partial G_y}{\partial y} \\ \cdot \\ \cdot \\ \frac{\partial G_y}{\partial y} \end{array} \right] \Delta y \quad (A2)$$

$$\Rightarrow G(k^0 + \Delta k) e(0) = G(k^0) e(0) + \Delta G(k^0) \text{diag}(e(0)) \Delta k \quad (A3)$$

And similarly for matrix $H$:

$$H_y(k^0 + \Delta k) = H_y(k^0) + \Delta H_y = \left[ \begin{array}{c} \frac{\partial H_y}{\partial y} \\ \cdot \\ \cdot \\ \frac{\partial H_y}{\partial y} \end{array} \right] \Delta y \quad (A4)$$

$$\Rightarrow H(k^0 + \Delta k) e = H(k^0) e + \Delta H(k^0) \text{diag}(e) \Delta k \quad (A5)$$

The matrix elements in (A2) and (A4) can be calculated by taking the partial derivatives to $k_i$ in expressions (8) and (9), while the integral of the latter equation can be discretized analogous to (16) for numerical calculation.

With the linearization of $G$ and $H$ in (A3) and (A5), the residue in definition (17) can be written in terms of a matrix $M$ working on $\Delta k$:

$$r(k^0 + \Delta k) = r^0 - M \Delta k,$$

with

$$M = G(k^0) \text{diag}(e(0)) + H(k^0) \text{diag}(e) \quad (A6)$$

Using these definitions, the cost function (A1) can be written in the quadratic form:

$$J(\Delta k) = \Delta k^T A_1 \Delta k - 2b_1^T \Delta k + c_1, \quad \text{with} \quad A_1 = M^T M, \quad b_1 = M^T r^0, \quad c_1 = r^0 r^T r^0 \quad (A7)$$

Because matrix $A_1$ is both symmetric and positive definite, $J$ has a global minimum at $\Delta k$ solved from $A_1 \Delta k = b_1$. However, the minimalization of the cost function is an ill-posed problem. To avoid unrealistic values and strong spatial fluctuations of the solution, we regularize the problem by demanding smoothness of the solution field. This can be accomplished by adding an extra constraint in the cost function (Tikhonov regularization):

$$J(k) = r(k)^T r(k) + k^T A_2 k \quad (A8)$$

where $A_2$ is composed of difference operators $D^x$ and $D^y$, which return all deviations from a linear interpolation of the nearest neighbors in the zonal and meridional direction of a solution field $k$:

$$d_i^x = k_i - \frac{1}{2} (k_{i+1} + k_{i-1}) \Rightarrow d^x = D^x k$$

$$d_i^y = k_i - \frac{1}{2} (k_{i+1} + k_{i-1}) \Rightarrow d^y = D^y k$$

$k_i$, $k_{i+1}$, $k_{i-1}$, $k^0$ are the neighboring values of element $k$ in western, eastern, northern, and southern direction, respectively. The regularization term in (A8) consists of a summation of all squared deviations in both directions, $d^x D^x + d^y D^y$, which means that the sparse matrix $A_2$ can be written as

$$A_2 = \alpha \left( D^T D^x + D^y D^y \right) \quad (A10)$$

An appropriate value of $\alpha$, which determines the relative weighting between the two constraints, is obtained by the L-curve method [Hansen, 1992].

The second term of (A8) can also be linearized for $\Delta k$ and written in the quadratic form, from which it can be shown that the minimum of the linearized cost function is found by solving $\Delta k$ from $A \Delta k = b$, with

$$A = M^T M + A_2, \quad b = M^T r^0 - A_2 k^0 \quad (A11)$$

This gives us a recipe for finding the minimum of the nonlinear cost function. We calculate $G$ and $H$ and their derivatives for an initial value $k^0$. Next we solve $\Delta k$ from $A \Delta k = b$ with the minimal residual method [Paige and Saunders, 1975], which gives us a new linearization point $k' = k^0 + \Delta k$. We iterate this procedure until the decrease in the cost function $J(k') - J(k'^{(i)})$ is less than 1% of $J(k')$.

Because the cost function is approximated by the quadratic form, the convergence rate will be of the second order.

Acknowledgments. A large part of this work was funded by the European Space Agency (ESA) in the Dragon-2 young scientist support and in the Global Emission project. The authors wish to thank Qiang Zhang for his help on the INTEX-B inventory, and Henk Eskes for giving valuable support and advice.

References


