

Numerical study of variational data assimilation algorithms based on decomposition methods in atmospheric chemistry models

Alexey Penenko^{1,2}, Pavel Antokhin³

¹Institute of Computational Mathematics and Mathematical Geophysics SB RAS, ICM&MG SB RAS, prospect Akademika Lavrentyeva 6, 630090, Novosibirsk, Russia

²Novosibirsk State University, NSU, Pirogova Str. 2, 630090, Novosibirsk, Russia

³V.E. Zuev Institute of Atmospheric Optics SB RAS, IAO SB RAS, Academician Zuev square 1, 634055, Tomsk, Russia

aleks@ommgp.sccc.ru

Abstract. The performance of a variational data assimilation algorithm for a transport and transformation model of atmospheric chemical composition is studied numerically in the case where the emission inventories are missing while there are additional in situ indirect concentration measurements. The algorithm is based on decomposition and splitting methods with a direct solution of the data assimilation problems at the splitting stages. This design allows avoiding iterative processes and working in real-time. In numerical experiments we study the sensitivity of data assimilation to measurement data quantity and quality.

1. Introduction

Forecasting of chemical weather is a challenge for researchers due to the nonlinear, multiscale and dynamic character of the corresponding transport and transformation processes in the atmosphere. It becomes even more challenging when the emission data are not exact. The aim of data assimilation algorithms is to improve the quality of model forecasts with measurement data arriving in the course of calculations. High dimensional atmospheric chemistry problems impose strict requirements on the numerical efficiency of the algorithms. We study the performance of an algorithm that is able to work without any iterative processes. It is achieved by using independent data assimilation procedures at the splitting stages. Overviews of chemical data assimilation can be found in [1, 2, 3].

We consider a scenario where the unknown emission data are estimated with indirect concentration measurements. The algorithm is used to reconstruct the emission effects of a substance with measurements of the other substances. We evaluate the sensitivity of data assimilation to the quality and quantity of measurement data.

2. Data assimilation algorithm

2.1. Transport and transformation model

Let us consider a horizontally homogeneous domain



$$z \in \Omega = [0, X_z], \quad t \in [0, T], \quad \Omega_T = \Omega \times [0, T],$$

bounded by $\partial\Omega_T = \partial\Omega \times [0, T]$. In the domain we consider the atmospheric chemistry transport and transformation model

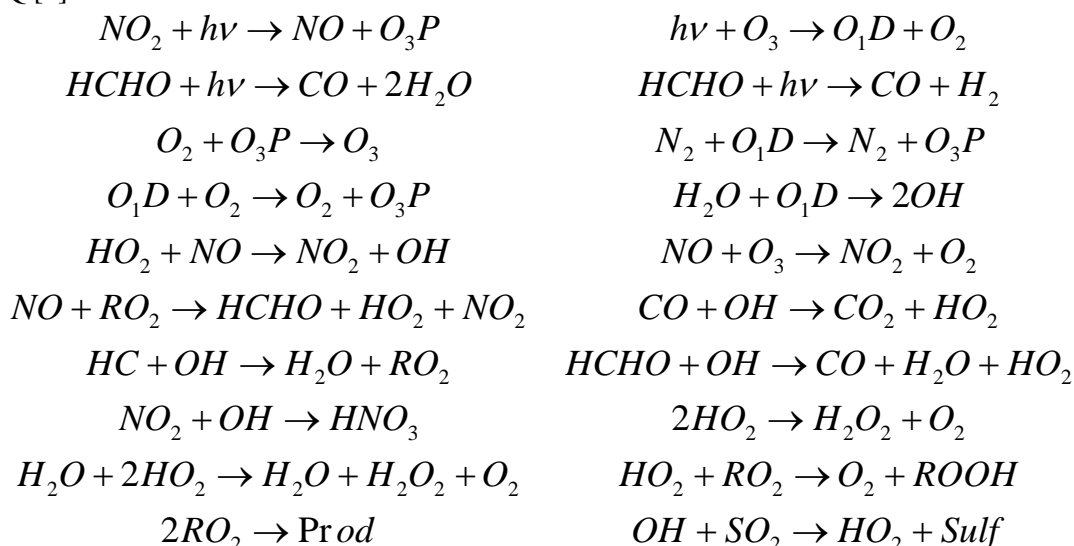
$$\frac{\partial \varphi_l(z, t)}{\partial t} + \text{div}(w \varphi_l(z, t) - \mu(z, t) \text{grad} \varphi_l(z, t)) \quad (1)$$

$$= S_l(\vec{\varphi}(z, t)) + f_l(z, t) + r_l(z, t), \quad (z, t) \in \Omega_T.$$

$$\mu(z, t) \frac{\partial \varphi_l(z, t)}{\partial \vec{n}} + \beta(z, t) \varphi_l(z, t) = g_l(z, t), \quad (z, t) \in \partial\Omega_T, \quad (2)$$

$$\varphi_l(z, 0) = \varphi_l^0(z), \quad z \in \Omega. \quad (3)$$

Here $\vec{\varphi}(z, t)$ is the state function. It has the physical meaning of the field of concentrations at point $(z, t) \in \Omega_T$, e.g. $\varphi_l(z, t)$ corresponds to the concentration of l^{th} substance at point (z, t) . Here $l = 1, \dots, N_c$, N_c is the total number of substances considered, $w(z, t)$ denotes the "wind speed", $\mu(z, t)$ is the diffusion coefficient, $S: \mathbb{R}^{N_c} \rightarrow \mathbb{R}^{N_c}$ is the transformation operator, \vec{n} is the boundary outer normal direction, $\vec{f}(z, t)$, $\vec{g}(z, t)$, $\vec{\varphi}^0(z)$ - a priori data from the sources and the initial data, $\vec{r}(z, t)$ is the control function (uncertainty). It is introduced in the perfect model structure to assimilate the data. Each entry of $\vec{f}(z, t)$, $\vec{g}(z, t)$, $\vec{r}(z, t)$, $\vec{\varphi}^0(z)$ vectors corresponds to a quantity attributed to l -th substance at point $(z, t) \in \Omega_T$. The transformation operator S is defined by a chemical kinetics system of 22 reacting species from [4,5] augmented with a SO_2 reaction taken from a model called CMAQ [6]:



The reaction rates are taken from [5]. They depend on time, i.e. photochemistry is considered. This kinetics system can be presented in production-destruction operator form:

$$S_l(\vec{\varphi}(z, t)) = -P_l(\vec{\varphi}(z, t))\varphi_l(z, t) + \Pi_l(\vec{\varphi}(z, t)), \quad l = 1, \dots, N_c \quad (4)$$

$$P_l, \Pi_l: \mathbb{R}_+^{N_c} \rightarrow \mathbb{R}_+, \quad (5)$$

where P_l is the destruction rate functional and Π_l is the production functional.

Direct problem: Given $\vec{f}(z, t)$, $\vec{g}(z, t)$, $\vec{r}(z, t)$, $\vec{\phi}^0(z)$, determine $\vec{\phi}(z, t)$ from equations (1)-(3). The exact solution $\vec{\phi}_*$ is the solution of the direct problem corresponding to the «unknown» emissions \vec{r}_* .

We assume all the functions and model parameters to be smooth enough for the solutions to exist and the transformations to be valid.

For the numerical solution, let us introduce a uniform temporal grid $\omega_t = \{t^j\}_{j=1}^{N_t}$ on $[0, T]$ with step size τ and N_t points and a uniform spatial grid ω_z with N_z , grid points on Ω . Let $Q(\omega)$ be the space of real grid functions on ω . The direct problem can be efficiently solved with a splitting method. Let us consider an additive-averaged splitting scheme (analogous to [7]) on the intervals $t^j \leq t \leq t^{j+1}$. The splitting is done with respect to the physical processes of advection-diffusion and transformation. Finally, we have two parallel stages with a step partition $\gamma_z + \gamma_c = 1$ and sources partition $\vec{f} = \vec{f}_c + \vec{f}_z$.

- The convection-diffusion process:

$$\gamma_z \frac{\partial \vec{\phi}_z(z, t)}{\partial t} + \frac{\partial}{\partial z} (w(z, t) \vec{\phi}_z(z, t)) - \frac{\partial}{\partial z} \left(\mu(z, t) \frac{\partial}{\partial z} \vec{\phi}_z(z, t) \right) = \vec{f}_z(z, t) + \vec{r}_z(z, t), \quad (z, t) \in \Omega \times [t^{j-1}, t^j], \quad (6)$$

$$\mu(z, t) \frac{\partial \vec{\phi}_\beta(z, t)}{\partial \vec{n}} + \beta(z, t) \vec{\phi}_z(z, t) = \vec{g}_a(z, t), \quad (z, t) \in \partial\Omega \times [t^{j-1}, t^j], \quad (7)$$

$$\vec{\phi}_\beta(z, t^{j-1}) = \vec{\phi}(z, t^{j-1}), \quad \vec{x} \in \Omega, \quad (8)$$

where $\partial\Omega = \{z = 0, z = X_z\}$. This initial value problem can be approximated in implicit matrix form:

$$\gamma_z \frac{\vec{\phi}_z^j - \vec{\phi}_z^{j-1}}{\tau} + L_z \vec{\phi}_z^j = \vec{r}_z^j + \vec{f}_z^j, \quad L_z(\vec{\phi}) := \left\{ L_z(\vec{\phi}_l) \right\}_{l=1}^{N_c}. \quad (9)$$

Here $\vec{\phi}^j \in Q(\omega)^{N_c}$ is the solution on the j -th time level, $\vec{r}^j \in Q(\omega)^{N_c}$ is the uncertainty on the j -th time level and $L_z : Q(\omega) \rightarrow Q(\omega)$ are the approximated advection-diffusion operators from equation (1) corresponding to the spatial dimensions.

- The chemical reaction processes:

$$\gamma_c \frac{\partial \vec{\phi}_c(z, t)}{\partial t} + \text{diag}(\vec{P}(\vec{\phi}_c(z, t))) \vec{\phi}_c(\vec{x}, t) = \vec{\Pi}(\vec{\phi}_c(z, t)) + \vec{f}_c(z, t) + \vec{r}_c(z, t), \quad (z, t) \in \Omega \times [t^{j-1}, t^j], \quad \vec{\phi}_c(z, t^j) = \vec{\phi}(z, t^j), \quad z \in \Omega,$$

Or, in entry form

$$\gamma_c \frac{\partial \phi_{cl}(z, t)}{\partial t} + P_l(\vec{\phi}_c(z, t)) \phi_{cl}(\vec{x}, t) = \Pi_l(\vec{\phi}_c(z, t)) + f_{cl}(z, t) + r_{cl}(z, t), \quad (z, t) \in \Omega \times [t^{j-1}, t^j], \quad l = 1, \dots, N_c.$$

In [8] a family of unconditionally monotonic schemes is constructed, from the first to fourth order of accuracy. One of the single stage schemes is equivalent to the well-known QSSA scheme [9]:

$$\vec{\phi}_c^j = \vec{G}(\vec{\phi}^{j-1}) + \text{diag}(\vec{W}(\vec{\phi}^{j-1})) \vec{r}_c^j. \quad (10)$$

where

$$G_l(\vec{\phi}^{j-1}(\bar{p})) = \phi_l^{j-1}(\bar{p}) e^{-P_l(\vec{\phi}^{j-1}(\bar{p}))\tau} + \frac{1 - e^{-P_l(\vec{\phi}^{j-1}(\bar{p}))\tau}}{P_l(\vec{\phi}^{j-1}(\bar{p}))\tau} \left(\Pi_l(\vec{\phi}^{j-1}(\bar{p})) + f_{cl}^j(\bar{p}) \right) \tau,$$

$$W_l(\vec{\phi}^{j-1}(\bar{p})) = \frac{1 - e^{-P_l(\vec{\phi}^{j-1}(\bar{p}))\tau}}{P_l(\vec{\phi}^{j-1}(\bar{p}))\tau} \tau.$$

The chemical splitting stage is calculated on a temporal grid $\hat{\omega}_t$ that is iNt times finer than the original temporal grid, i.e. every interval $[t^j, t^{j+1}]$ is divided to iNt subintervals.

- The next step approximation is

$$\vec{\phi}^j = \vec{\phi}_z^j + \vec{\phi}_c^j. \quad (11)$$

An advantage of the scheme is that at each time step the individual processes are evaluated independently (in parallel).

2.2. Weakly coupled data assimilation algorithm

In order to assimilate the measurement data, we connect the measured quantities with the model variables. This is formally done with a measurement operator H :

$$\vec{I}(t) = H(t, \vec{\phi}^*(., t)) + \vec{\eta}(t), \quad t \in [0, T], \quad (12)$$

where $\vec{I}(t)$ is the measurement data, $\vec{\phi}^*(., t)$ is the «true» (or exact) solution, $\vec{\eta}(t)$ is the measurement data uncertainty.

Data assimilation problem: Determine $\vec{\phi}^*(., t)$ for $t > t^*$ with equations (1)-(3), (12) and functions \vec{f}_a , \vec{g}_a , $\vec{\phi}_a^0$, \vec{I} defined on $0 < t \leq t^*$.

We consider N_M *in situ* measurements at the domain grid points $\{(z_M^m, t_M^m)\}_{m=1}^{N_M} \subset \omega \times \omega_t$. Hence the m -th measurement is defined by the vector

$$\xi_m = \{(z_M^m, t_M^m, I_m^m, I_m, \sigma_M^m)\}, \quad m = 1, \dots, N_M,$$

where z_M^m is the measurement point location, t_M^m is the measurement time point, I_M^m is the number of substances being measured, I_m is the resulting concentration and σ_M^m is the standard variation of the measurement. According to the data assimilation problem statement, at time t^j we can use only the measurements with $t_M^m \leq t^j$. Let us define a set of indices

$$\theta^j = \{1 \leq m \leq N_M \mid t_M^m \leq t^j\}.$$

The corresponding measurement operator and measurement data are defined by

$$H^j \vec{\phi} = \left\{ \phi_{I_m^m}^j(z_M^m, t_M^m) \right\}_{m \in \theta^j}, \quad I^j = \left\{ \left(1 + (\sigma_M^m)^2 \right) \xi \right\}_{m \in \theta^j} \phi_{I_m^m}^*(z_M^m, t_M^m), \quad \sigma^j = \{ \sigma_M^m \}_{m \in \theta^j}.$$

We consider regular measurements in space and time which are defined by temporal Δt_m and spatial Δz_m gaps between the measurement points, ξ is a normally distributed random variable.

The variational data assimilation method provides a solution to the data assimilation problem as the minimum of a functional with constraints imposed by the model. The functional usually combines the measurement data misfit with the norm of the control variable:

$$J^j(\vec{\phi}, \vec{r}) = \alpha \|H^j \vec{\phi} - \vec{I}^j\|_{\sigma_j}^2 + \|\vec{r}\|^2,$$

where $\|\cdot\|$ is the norm of a Hilbert space over $Q(\omega)^{N_c}$ and $\langle \cdot, \cdot \rangle$ is the corresponding inner product, α is a regularization (assimilation) parameter which is selected to make the solution closer to the direct model solution or to the measurements. At t^j we update only the control variable \vec{r}^j for this time step. Under a weakly-coupled (or fine-grained) approach [10, 11, 12, 13] the same data are assimilated at the different splitting stages and the results are coupled afterwards. We seek for a minimum of the functional

$$J_f^j\left(\left\{\vec{\phi}_\beta^j, \vec{r}_\beta^j\right\}_{\beta \in \{z, c\}}\right) = \sum_{\beta \in \{z, c\}} J^j\left(\vec{\phi}_\beta^j, \vec{r}_\beta^j\right)$$

under the constraints defined by equations (9), (10) with respect to the independent \vec{r}_β . Using the method of Lagrange multipliers to solve minimization problems with equality constraints, we can construct the augmented functional

$$\begin{aligned} \bar{J}_f\left(\left\{\vec{\phi}_\beta^j, \vec{r}_\beta^j\right\}_{\beta \in \{z, c\}}\right) &= \sum_{\beta \in \{z, c\}} J^j\left(\vec{\phi}_\beta^j, \vec{r}_\beta^j\right) + \left\langle \gamma_z \frac{\vec{\phi}_z^j - \vec{\phi}^{j-1}}{\tau} - L_z \vec{\phi}^j - \vec{r}_z^j - \vec{f}_z^j, \vec{\psi}_z^j \right\rangle \\ &+ \left\langle \vec{\phi}_c^j - \vec{G}(\vec{\phi}^{j-1}) - \text{diag}(\vec{W}(\vec{\phi}^{j-1})) \vec{r}_c^j, \vec{\psi}_c^j \right\rangle. \end{aligned}$$

The components of $\bar{J}_f\left(\left\{\vec{\phi}_\beta^j, \vec{r}_\beta^j\right\}_{\beta \in \{z, c\}}\right)$ corresponding to the different β are independent, hence the stationary point coordinates can be found independently.

In order to present an algorithm of finding a stationary point for the convection-diffusion stage, we need further elaboration of the operator L . We use approximations of equation (6) that produce the tridiagonal matrix systems

$$-a_i \phi_{i+1}^j + b_i \phi_i^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j, i = 0, \quad (13)$$

$$-a_i \phi_{i+1}^j + b_i \phi_i^j - c_i \phi_{i-1}^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j, i = 1, \dots, N-1, \quad (14)$$

$$b_i \phi_i^j - c_i \phi_{i-1}^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j, i = N. \quad (15)$$

Here the assimilated state is the solution of the minimization problem

$$J(\phi^j, r^j) \tau = \left(\alpha \sum_{i=0}^{N_z} \left(\frac{\phi_i^j - I_i^j}{\sigma_i} \right)^2 M_i^j + \sum_{i=0}^{N_z} (r_i^j)^2 \right) \tau,$$

under the constraints defined by equations (13)-(15) where M_i^j is a spatial-temporal measurement mask (i.e. M_i^j is equal to 1 if there is a measurement at point (z_i, t^j) and to 0 otherwise), I_i^j is the measurement data at point (z_i, t^j) (if there is a measurement) and σ_i is the measurement device standard deviation of the measurement at point (z_i, t^j) (if there is a measurement). Introducing Lagrange multipliers, we obtain the augmented functional

$$\bar{J}_f(\phi^j, r^j, \psi^j) \tau = J(\phi^j, r^j) \tau + \sum_{i=0}^N \left(-a_i \phi_{i+1}^j + b_i \phi_i^j - c_i \phi_{i-1}^j - \phi_i^{j-1} - \tau r_i^j - \tau f_i^j \right) \psi_i^j.$$

Equating the first variations of the augmented functional to zero, we obtain the following tridiagonal matrix equation [10, 11, 12]:

$$\begin{aligned} -A_i \Phi_{i+1}^j + B_i \Phi_i^j &= F_i^j, i = 0, \\ -A_i \Phi_{i+1}^j + B_i \Phi_i^j - C_i \Phi_{i-1}^j &= F_i^j, i = 1, \dots, N-1, \\ B_i \Phi_i^j - C_i \Phi_{i-1}^j &= F_i^j, i = N, \end{aligned}$$

$$A_i = \begin{pmatrix} a_i & 0 \\ 0 & c_{i+1} \end{pmatrix}, B_i = \begin{pmatrix} b_i & -\frac{\tau}{2} \\ \frac{2\alpha M_i \tau}{\sigma_i^2} & b_i \end{pmatrix}, C_i = \begin{pmatrix} c_i & 0 \\ 0 & a_{i-1} \end{pmatrix}, \Phi_i^j = \begin{pmatrix} \phi_i^j \\ \psi_i^j \end{pmatrix}, F_i^{j+1} = \begin{pmatrix} \phi_i^{j-1} + \tau f_i^j \\ \frac{2\alpha M_i \tau}{\sigma_i^2} I_i^j \end{pmatrix},$$

which can be solved by a direct Gaussian elimination method.

At the transformation stage the algorithm is the same for any grid point $\bar{p} \in \omega$. For brevity let $\vec{\phi}^j = \vec{\phi}_c^j(\bar{p}) \in \mathbb{R}^{N_c}$, $\vec{r}^j = \vec{r}_c^j(\bar{p}) \in \mathbb{R}^{N_c}$, $\vec{\psi}^j = \vec{\psi}_c^j(\bar{p}) \in \mathbb{R}^{N_c}$ and the result is sought-for as a stationary point of the augmented functional

$$\bar{J}(\vec{\phi}^j, \vec{r}^j) = \alpha \sum_{l=1}^{N_c} \left(\frac{\phi_l^j - I_l^j}{\sigma_l} \right)^2 M_l^j + \sum_{l=1}^{N_c} (r_l^j)^2 + \sum_{l=1}^{N_c} \left(\phi_l^j - G_l(\vec{\phi}^{j-1}(\bar{p})) - W_l(\vec{\phi}^{j-1}(\bar{p})) r_l^j \right) \psi_l^j,$$

where M_l^j is equal to 1 if the l -th substance is measured at point \bar{p} at moment t^j and zero otherwise. This minimum is given by the interpolation formula

$$\phi_l^j = \frac{1}{1 + Z_l^j} G_l(\vec{\phi}^{j-1}(\bar{p})) + \frac{Z_l^j}{1 + Z_l^j} I_l^j, \quad Z_l^j = \frac{\alpha M_l^j}{\sigma_l^2} W_l^2(\vec{\phi}^{j-1}(\bar{p})).$$

The next step is evaluated as in the direct problem:

$$\vec{\phi}^j = \vec{\phi}_z^j + \vec{\phi}_c^j.$$

Note that the resulting algorithm can be implemented without iterations. This allows using it in real-time for large chemical transformation models.

3. Numerical experiments

In the numerical experiments, we take a realistic scenario corresponding to August 8, 2013. The turbulence diffusion coefficient $\mu(z, t)$ is calculated from ground heat flux measurements and the friction stress, and from vertical wind profiles based on the k -theory. The vertical speed $w(z, t)$ and $\mu(z, t)$ are presented in figure 1. Realistic initial conditions are considered. The grid parameters are $N_z = 30, X_z = 3000m, T = 75600s, N_t = 281, iNt = 10^4$.

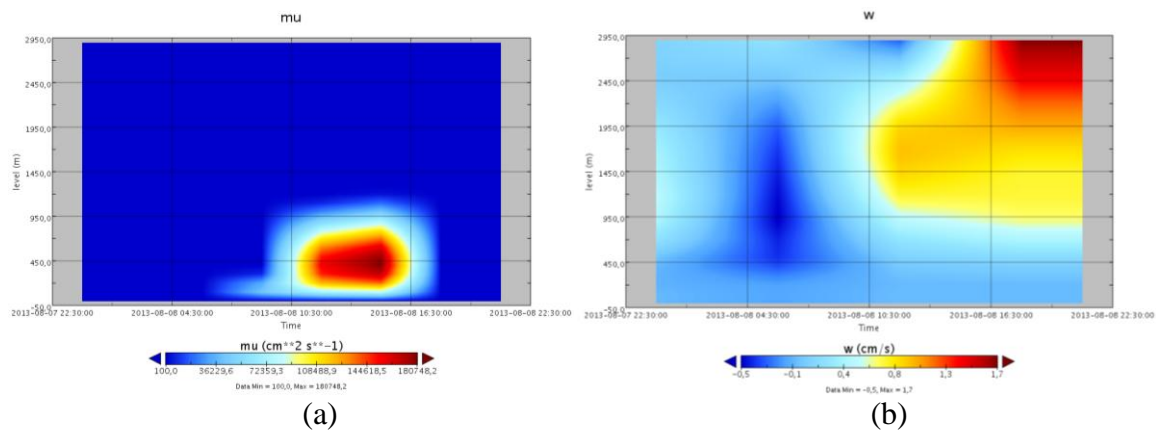


Figure 1. Turbulent diffusion coefficient $\mu(z, t)$ (a) and vertical speed $w(z, t)$ (b).

We compare three configurations for the numerical experiments:

- NoDA (Background solution): Direct problem solution that is driven by initial and boundary conditions only.
- «Truth» (Exact solution): Direct problem solution that is driven by both initial and boundary conditions and by an «unknown» source of NO_2 «on the ground» ($z=0$). The emission rate is chosen to provide a 100% relative difference between «true» and background solutions. The scenario models traffic emissions.
- DA (Data assimilation solution): The result of data assimilation algorithm. In this case, the initial and boundary conditions are the same as for NoDA and «true» solutions. The source is absent as in the NoDA case but there are NO and O_3 concentration measurements available for the data assimilation system. In the experiments $\sigma_M^m = 1$, $\alpha = 10^{10}$.

The results are presented for three groups of substances: a substance whose measurement data is assimilated (figure 2), a substance that is emitted in the «true» solution (figure 3) and a substance that is neither emitted nor assimilated (figure 4). In the experiment $\sigma = 0$, $\Delta z_m = 500$ m, $\Delta t_m = 135$ min.

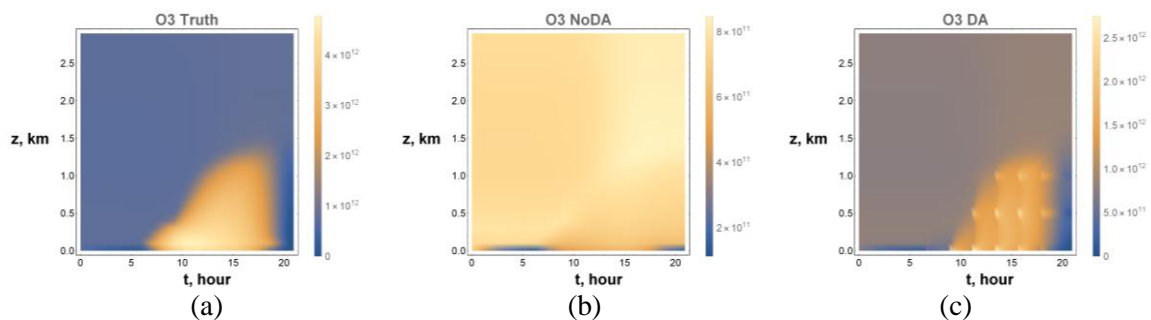


Figure 2. Comparison of «true» solution (a), solution without data assimilation (b) and data assimilation results (c) for a substance whose concentrations are assimilated.

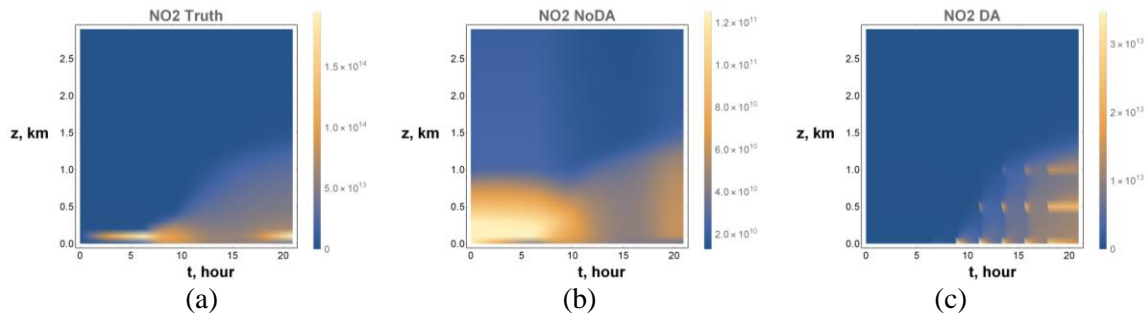


Figure 3. Comparison of «true» solution (a), solution without data assimilation (b) and data assimilation results (c) for a substance that is emitted.

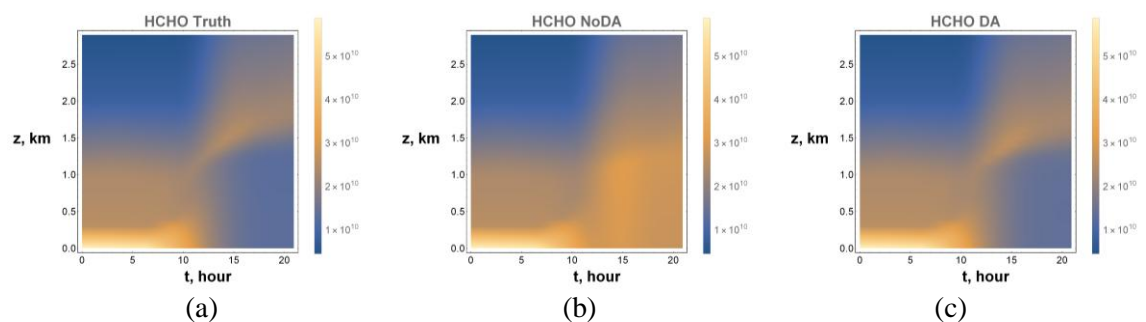


Figure 4. Comparison of «true» solution (a), solution without data assimilation (b) and data assimilation results (c) for a substance that is neither assimilated nor emitted.

From these figures we can conclude that for these substances the data assimilation algorithm was successful at least qualitatively. In the following experiments we will study the sensitivity of the results to measurements data. The results for the substances affected by the emissions (NoDA relative error $> 0.5\%$) are presented in tables 1-3.

Experiment 1: Sensitivity to vertical gaps in data. Δz_m is varied with a fixed $\sigma = 0$, $\Delta t_m = 269$ min .

Table 1. Relative error of DA versus vertical gaps in measurement points.

Vertical gap Δz_m	H2O2	HC	HCHO	HNO3	HO2	NO	NO2
500 m	0.01	0.01	0.07	0.63	0.25	0.79	0.95
1000 m	0.01	0.01	0.07	0.65	0.26	0.83	0.96
1500 m	0.01	0.01	0.08	0.66	0.27	0.84	0.97
NoDA	0.01	0.03	0.32	0.76	0.53	0.99	1

Vertical gap Δz_m	O1D	O3	O3P	OH	RO2	ROOH	Prod
500 m	0.63	0.59	0.93	0.3	0.27	0.22	0.22
1000 m	0.66	0.61	0.94	0.31	0.28	0.24	0.23
1500 m	0.66	0.62	0.95	0.33	0.29	0.25	0.24
NoDA	0.76	0.7	1	0.87	0.57	0.5	0.48

Experiment 2: Sensitivity to temporal gaps in data. Δt_m is varied with a fixed $\sigma = 0$, $\Delta z_m = 500$ m .

Table 2. Relative error of DA versus temporal gaps in measurements.

Temporal gap, Δt_m	H2O2	HC	HCHO	HNO3	HO2	NO	NO2
135min	0	0	0.05	0.55	0.18	0.69	0.9
269 min	0.01	0.01	0.07	0.63	0.25	0.79	0.95
404 min	0.01	0.01	0.14	0.69	0.35	0.88	0.97
NoDA	0.01	0.03	0.32	0.76	0.53	0.99	1

Temporal gap, Δt_m	O1D	O3	O3P	OH	RO2	ROOH	Prod
135 min	0.55	0.52	0.87	0.22	0.2	0.15	0.15
269 min	0.63	0.59	0.93	0.3	0.27	0.22	0.22
404 min	0.69	0.64	0.96	0.5	0.38	0.31	0.3
NoDA	0.76	0.7	1	0.87	0.57	0.5	0.48

Experiment 3: Sensitivity to measurement noise. σ is varied with a fixed $\Delta t_m = 269$ min ,
 $\Delta z_m = 500$ m .

Table 3. Relative error of DA versus standard deviation of measurement noise.

Measurement noise, σ	H2O2	HC	HCHO	HNO3	HO2	NO	NO2
0	0.01	0.01	0.07	0.63	0.25	0.79	0.95
0.1	0.01	0.01	0.07	0.64	0.26	0.81	0.95
0.5	0.01	0.01	0.09	0.67	0.29	0.85	0.96
1	0.01	0.01	0.12	0.71	0.34	0.9	0.98
2	0.01	0.02	0.2	0.74	0.42	0.94	1.04
3	0.01	0.02	0.2	0.75	0.41	0.96	1.22
NoDA	0.01	0.03	0.32	0.76	0.53	0.99	1

Measurement noise, σ	O1D	O3	O3P	OH	RO2	ROOH	Prod
0	0.63	0.59	0.93	0.3	0.27	0.22	0.22
0.1	0.64	0.6	0.93	0.31	0.28	0.24	0.23
0.5	0.68	0.63	0.95	0.36	0.31	0.28	0.27
1	0.72	0.67	0.97	0.45	0.37	0.34	0.33
2	0.75	0.71	1	0.67	0.45	0.41	0.39
3	0.76	0.74	1.03	0.67	0.44	0.42	0.4
NoDA	0.76	0.7	1	0.87	0.57	0.5	0.48

The quantitative efficiency of the data assimilation can be studied by comparison of the NoDA results and the results of data assimilation (DA). The small error in both DA and NoDA cases for a substance means that the emissions do not affect the dynamics of its concentrations (as it is for HC and H₂O₂). The less error for the DA cases compared to NoDA means that the data assimilation

is efficient. In the numerical experiments with high noise levels ($\sigma = 2, 3$), the DA becomes inefficient for some substances. In all DA cases the more (smaller gaps) and the better (less noise) data are available the better are the reconstructions.

4. Conclusions

The combination of splitting and direct variational data assimilation schemes at the splitting stages allows constructing computationally efficient algorithms for data assimilation of *in situ* concentration measurements in the convection-diffusion-reaction models. The algorithm performance was evaluated in a series of numerical experiments with indirect measurements. In the numerical experiments, the data assimilation algorithm was efficient in all the experiments except for ones with large noise levels. For all the data assimilation cases, the smaller gaps and less noise in data, the better the reconstruction results. We conclude that in the numerical experiments the data assimilation algorithm was able to partially substitute the missing emission inventories with additional concentration measurements.

Acknowledgments

The work is partially supported by the Presidium of RAS under the Programs I.33P and II.2P/I.3-3; by the projects MK-8214.2016.1 and RFBR 14-01-00125.

References

- [1] Sandu A, Chai T 2011 Chemical data assimilation - an overview *Atmos.* **2** 426–63
- [2] Bocquet M, Elbern H, Eskes H et al. 2015 Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models *Atmos. Chem. Phys.* **15** 5325–58.
- [3] Elbern H, Strunk A, Schmidt H, Talagrand O 2007 Emission rate and chemical state estimation by 4-dimensional variational inversion *Atmos. Chem. Phys.* **7** 3749–69
- [4] Stockwell W, Goliff W 2002 Comment on Simulation of a reacting pollutant puff using an adaptive grid algorithm by R. K. Srivastava et al *J. Geophys. Res.* **D.107** 4643–50.
- [5] Gery M W, Whitten G Z, Killus J P, Dodge M C 1989 A photochemical kinetics mechanism for urban and regional scale computer modeling *J. Geophys. Res.* **94** 12952–6
- [6] Byun D W, Schere K L 2006 Review of the governing equations, computational algorithms and other components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System *Appl. Mech. Rev.* **59**(2) 51–77
- [7] Samarskii A A, Vabishchevich P N 1995 *Computational Heat Transfer Vol.1,2* (Chichester :Wiley) 1–850
- [8] Penenko V, Tsvetova E 2013 Variational methods for construction of monotone approximations for atmospheric chemistry models *Numer. Anal. Appl.* **6**(3) 210–20
- [9] Hesstvedt E, Hov O, Isaacsen I 1978 Quasi-steady-state-approximation in air pollution modelling: comparison of two numerical schemes for oxidant prediction *Int. J. Chem. Kinet.* **10** 971–94
- [10] Penenko A 2006 Some theoretical and applied aspects of sequential variational data assimilation *Comp. tech.* **11**(2) 35–40
- [11] Penenko V V 2009 Variational methods of data assimilation and inverse problems for studying the atmosphere, ocean, and environment *Num. Anal. and Appl.* **2**(4) 341–51
- [12] Penenko A V, Penenko V V 2014 Direct data assimilation method for convection-diffusion models based on splitting scheme *Comp. tech.* **19**(4) 69–83
- [13] Penenko A et al. 2015 Direct variational data assimilation algorithm for atmospheric chemistry data with transport and transformation model *Proc. SPIE* **9680** 968076–1–12