Variational data assimilation schemes for transport and transformation models of atmospheric chemistry

Alexey V. Penenko[1,2], Vladimir V. Penenko[1,2], Elena A. Tsvetova[1], Pavel N. Antokhin[3]

[1]Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

[2]Novosibirsk State University, Novosibirsk, Russia

[3]V.E. Zuev Insititute of Atmopsheric Optics, Tomsk, Russia

aleks@ommgp.sscc.ru, penenko@sscc.ru, e.tsvetova@ommgp.sscc.ru, apn@iao.ru

Introduction

- Atmospheric chemistry dynamics is studied with convection-diffusion-reaction model. Measurement data is provided as the point-wise concentration measurements as it can be done by an automated city monitoring network.
- The numerical Data Assimilation (DA) algorithm presented is based on the additive-averaged splitting scheme. It carries out "fine-grained" variational data assimilation on the separate splitting stages.
- This design provides efficient implementation due to the data assimilation scheme applied to the separate transport process on each coordinate line containing in situ measurement points.
- The goal of the work is to investigate the performance of the algorithm.

Data assimilation problem for transport and transformation model

- 4D models describing the processes of heat, moisture, radiation, and pollutants transport and transformation in the atmosphere have the generic structure:

$$\frac{\partial \rho\vec{\phi}}{\partial t} + \mathrm{div}\rho(\vec{\phi}\vec{u} - \mu\mathrm{grad}\,\vec{\phi}) + \rho(S\vec{\phi}) - \vec{f}_a - \vec{r}) = 0,$$

$$\vec{\phi}^0 = \vec{\phi}_a^0,\ R_{bound}(\vec{\phi}) = \vec{g}_a.$$

  Here $\vec{\phi}(x,t)$ are species concentrations vector, $\{\rho, \vec{u}, \mu\}$ are model parameters that can be calculated with a hydrodynamics model, S is transformation operator, $\vec{f}_a$, $\vec{g}_a$, $\vec{\phi}_a^0$ are *a priori* values of sources and initial data.

- Incoming measurement data is connected with the state function by means of observation operator $\mathbf{H}$

$$\vec{I}_m = [\mathbf{H}(\vec{\phi})]_m + \vec{\eta},$$

- Flexibility (uncertainty/control) functions $\vec{r}$, $\vec{\eta}$ are introduced in the "perfect" model structure. $\vec{\eta}$ is constrained by a measurement noise specification (e.g. its norm).
- **Data assimilation problem:** State $\vec{\phi}$ has to be estimated at $t > t_*$ with measurement data for $t < t_*$.

Split model

Consider additive-averaged splitting scheme [1] on time interval $t^j \le t \le t^{j+1}$:

- Transformation process:

$$\gamma_c \frac{\partial \vec{\phi}_c}{\partial t} = S(\vec{\phi}_c) + \vec{r}_c, \quad \vec{\phi}_c\left(t^j\right) = \vec{\phi}\left(t^j\right).$$

- Transport processes ($\alpha = \overline{1,3}$):

$$\gamma_\alpha \frac{\partial \vec{\phi}_\alpha}{\partial t} + L_\alpha\,\vec{\phi}_\alpha = \vec{f}_\alpha + \vec{r}_\alpha, \quad \vec{\phi}_\alpha\left(t^j\right) = \vec{\phi}\left(t^j\right).$$

- Next step approximation:

$$\vec{\phi}(t^{j+1}) = \sum_{\alpha=1}^{3} \gamma_\alpha \vec{\phi}_\alpha(t^{j+1}) + \gamma_c \vec{\phi}_c(t^{j+1}), \quad \sum_{\alpha=1}^{3} \gamma_\alpha + \gamma_c = 1.$$

An atmospheric chemistry model

Consider a system of 22 reacting species from [2] augmented with an SO_2 reaction taken from CMAQ model.

hv + NO2 → NO + O3P	hv + O3 → O1D + O2
HCHO + hv → CO + 2HO2	HCHO + hv → CO + H2
O2 + O3P → O3	N2 + O1D → N2 + O3P
O1D + O2 → O2 + O3P	H2O + O1D → 2OH
HO2 + NO → NO2 + OH	NO + O3 → NO2 + O2
NO + RO2 → HCHO + HO2 + NO2	CO + OH → CO2 + HO2
HC + OH → H2O + RO2	HCHO + OH → CO + H2O + HO2
NO2 + OH → HNO3	2HO2 → H2O2 + O2
H2O + 2HO2 → H2O + H2O2 + O2	HO2 + RO2 → O2 + ROOH
2RO2 → Prod	OH + SO2 → HO2 + Sulf

Reaction rates have been taken from [2] and are dependent on time i.e.photochemistry is considered.

References

[1] A.A. Samarskii and P.N. Vabishchevich. Computational Heat Transfer. Vol.1,2. Wiley, Chichester, 1995.

[2] William R. Stockwell and Wendy S. Goliff *Comment on "Simulation of a reacting pollutant puff using an adaptive grid algorithm" by R. K. Srivastava et al.* JOURNAL OF GEOPHYSICAL RESEARCH, 2002 V 107 No D22.

[3] V.V. Penenko, E.A. Tsvetova, A.V. Penenko. Variational approach and Euler's integrating factors for environmental studies// Computers and Mathematics with Applications, (2014) V.67, Issue 12, P. 2240Ü2256.

[4] Hesstvedt, E.; Hov, O.; Isaacsen, I. Quasi-steady-state-approximation in air pollution modelling: comparison of two numerical schemes for oxidant prediction. Int. J. Chem. Kinet. 1978, 10, 971-994.

[5] Penenko, A.V. and Penenko V.V. (2014). Direct data assimilation method for convection-diffusion models based on splitting scheme. Computational technologies, 19 Issue 4, 69-83 (In Russian).

[6] V. V. Penenko *Variational methods of data assimilation and inverse problems for studying the atmosphere, ocean, and environment* Num. Anal. and Appl., 2009 V 2 No 4, 341-351.

[7] A. Penenko *Some theoretical and applied aspects of sequential variational data assimilation (In Russian)*, Comp. tech. v.11, Part 2, (2006) 35-40.

[8] Nuterman R. et. al *High resolution forecast meteorology and chemistry for a Danish Urban Area* EMS 2011.

[9] Penenko A., Penenko V., Nuterman R., et al. Direct variational data assimilation algorithm for atmospheric chemistry data with transport and transformation model //Proc. SPIE 9680, 21st International Symposium Atmospheric and Ocean Optics: Atmospheric Physics, 968076 (November 19, 2015); doi:10.1117/12.2206008

[10] V. V. Penenko, E. A. Tsvetova, and A. V. Penenko Methods Based on the Joint Use of Models and Observational Data in the Frameworkof Variational Approach to Forecasting Weather and Atmospheric Composition Quality // Russian meterology and hydrology Vol.40 (6) p.365-373

[11] V. V. Penenko, E. A. Tsvetova, and A.V. Penenko Development of Variational Approach for Direct and Inverse Problems of Atmospheric Hydrodynamics and Chemistry // Izvestiya, Atmospheric and Oceanic Physics vol.51 (3) p.311-319

[12] Baklanov A., U. Korsholm, A. Mahura, C. Petersen, A. Gross, 2008: Enviro-HIRLAM: on-line coupled modelling of urban meteorology and air pollution. Adv. Sci. Res., 2, 41-46

[13] Flemming, J., A., Inness, H., Flentje, V., Huijnen, P., Moinat, M. G., Schultz, and Stein, O.: Coupling global chemistry transport models to ECMWF's integrated forecast system, Geosci. Model Dev., 2, 253-265, doi:10.5194/gmd-2-253-2009, 2009.

[14] AirBase - The European air quality database. http://www.eea.europa.eu/data-and-maps

Sequential variational emission rate reconstruction

Consider a discretized "truth" model:

$$\bar{\varphi}^0 = \bar{\varphi}_0, \tag{1}$$

$$L^j\,\bar{\varphi}^j = \bar{\varphi}^{j-1} + \tau f^j + \tau\bar{r}^j, \quad j = 1,...,\Theta, \tag{2}$$

with measurement data

$$H^j\bar{\varphi}^j = I^j + \delta I^j, \quad \delta\bar{I}^j = \left\|\delta I^j\right\|_Y, \quad j = 1,...,\Theta. \tag{3}$$

The data assimilation problem is a sequence of inverse problems. Here the j-th inverse problem is: with $\{L^m, f^m, m = 1,...,\Theta\}$ and $\{H^m, I^m, \delta^m, m = 1,...,j\}$ find $\{\bar{\varphi}^m, j = 1,...,\Theta\}$.
Consider a sequential variational data assimilation algorithm

$$\{\varphi^j, r^j\} = \underset{\{\varphi,r\}\in Q^j}{\arg\min}\, J^j(\varphi, r), \tag{4}$$

$$J^j(\varphi, r) = \left\|H^j\varphi - I^j\right\|^2 + \alpha\left\|r\right\|^2, \quad Q^j = \{\{\varphi, r\}, L^j\varphi = F^j + \tau r\}, \tag{5}$$

or

$$\begin{bmatrix} L^j & 0 & -\tau \\ 2H^{j*}H^j & \frac{1}{\tau}L^{j*} & 0 \\ 0 & -1 & 2\alpha \end{bmatrix} \begin{bmatrix} \varphi^j \\ \varphi^{j*} \\ r^j \end{bmatrix} = \begin{bmatrix} F^j \\ 2H^{j*}I^j \\ 0 \end{bmatrix}. \tag{6}$$

Exact solution of a sequential data assimilation problem is:

$$\tau\hat{r}^j := \left(\bar{\varphi}^{j-1} - \varphi^{j-1}\right) + \tau\bar{r}^j. \tag{7}$$

then

$$\tau\hat{r}^j - \tau r^j = \sum_{k=1}^{M} v_k \left(\frac{\alpha}{\alpha + s_k^2}\left\langle v_k, \tau\hat{r}^j\right\rangle_X + \frac{s_k^2}{\alpha + s_k^2}\frac{\left\langle u_k, \tau\delta I^j\right\rangle_Y}{s_k}\right) + \tau\hat{r}_\parallel^j,$$

where $\{v_k\}_{k=1,...,M}$ and $\{u_k\}_{k=1,...,M}$ are left and right singular vectors of $S = \tau HL^{-1}$, $\{s_k\}_{k=1,...,M}$ are s-numbers $\tau\hat{r}_\parallel^j = \tau\hat{r}^j - \sum_{k=1}^{M}\left\langle v_k, \tau\hat{r}^j\right\rangle_X v_k$.
Assimilation parameter α can be chosen with the discrepancy principle:

- Let a background forecast be $L\varphi_0^j = F^j$.
- If $\left\|I^j - H\varphi_0^j\right\|^2 \le \delta^2$ then $\varphi^j = \varphi_0^j$ otherwise find $\alpha^*(\delta)$ such that $\left\|I^j - H\varphi_{\alpha^*(\delta)}^j\right\|^2 = \delta^2$.

Direct algorithm for solution of (6) allows to produce a computationally effective algorithm. In this case according to Tikhonov $\lim_{\delta\to 0} r_{\alpha^*(\delta)} = \bar{r}$, where $\bar{r}$ is the normal solution of

$$\tau HL^{-1}r = I^j - HL^{-1}F^j = HL^{-1}\left(\bar{\varphi}^{j-1} - \varphi^{j-1} + \tau\bar{r}^j\right). \tag{8}$$

The problem is to solve (6).

One-dimensional direct data assimilation algorithm for convection-diffusion

In case of tridiagonal matrix process models

$$-a_i\phi_{i+1}^j + b_i\phi_i^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j, i = 0, \tag{9}$$

$$-a_i\phi_{i+1}^j + b_i\phi_i^j - c_i\phi_{i-1}^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j,\ i = 1,\ldots,N-2, \tag{10}$$

$$b_i\phi_i^j - c_i\phi_{i-1}^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j,\ i = N-1, \tag{11}$$

and in situ concentration measurements the solution of the minimization problem

$$\Phi(\phi^j, r^j) = \left(\sum_{i=1}^{N-1}\left(\frac{\phi_i^j - I_i^j}{\sigma_i}\right)^2 M_i^j + \alpha\sum_{i=1}^{N-1}\left(r_i^j\right)^2\right)\frac{\tau}{2},$$

WRT (9)-(11) where M_i^{j+1} is the spatial-temporal measurement mask and σ_i are measurement device standard deviations. Introducing Lagrange multipliers we obtain augmented functional with its stable point solving the tridiagonal matrix equation [5–7]

$$-\begin{pmatrix} a_i & 0 \\ 0 & c_{i+1} \end{pmatrix}\begin{pmatrix} \phi_{i+1}^j \\ \phi_{i+1}^{*j} \end{pmatrix} + \begin{pmatrix} b_i & -\frac{\tau}{\alpha} \\ \frac{M_i\tau}{\sigma_i^2} & b_i \end{pmatrix}\begin{pmatrix} \phi_i^j \\ \phi_i^{*j} \end{pmatrix} = \begin{pmatrix} \phi_i^{j-1} + \tau f_i^j \\ \frac{M_i\tau}{\sigma_i^2} I_i^j \end{pmatrix},$$

$$-\begin{pmatrix} a_i & 0 \\ 0 & c_{i+1} \end{pmatrix}\begin{pmatrix} \phi_{i+1}^j \\ \phi_{i+1}^{*j} \end{pmatrix} + \begin{pmatrix} b_i & -\frac{\tau}{\alpha} \\ \frac{M_i\tau}{\sigma_i^2} & b_i \end{pmatrix}\begin{pmatrix} \phi_i^j \\ \phi_i^{*j} \end{pmatrix} - \begin{pmatrix} c_i & 0 \\ 0 & a_{i-1} \end{pmatrix}\begin{pmatrix} \phi_{i-1}^j \\ \phi_{i-1}^{*j} \end{pmatrix} = \begin{pmatrix} \phi_i^{j-1} + \tau f_i^j \\ \frac{M_i\tau}{\sigma_i^2} I_i^j \end{pmatrix},$$

$$\begin{pmatrix} b_i & -\frac{\tau}{\alpha} \\ \frac{M_i\tau}{\sigma_i^2} & b_i \end{pmatrix}\begin{pmatrix} \phi_i^j \\ \phi_i^{*j} \end{pmatrix} - \begin{pmatrix} c_i & 0 \\ 0 & a_{i-1} \end{pmatrix}\begin{pmatrix} \phi_{i-1}^j \\ \phi_{i-1}^{*j} \end{pmatrix} = \begin{pmatrix} \phi_i^{j-1} + \tau f_i^j \\ \frac{M_i\tau}{\sigma_i^2} I_i^{j+1} \end{pmatrix}.$$

which is solved by the matrix sweep method.

Explicit chemistry data assimilation algorithm

Chemical kinetic equations defining transformation operator have the following structure

$$\frac{\partial\phi_l}{\partial t} + P_l(\bar{\phi})\phi_l = \Pi_l(\phi) + r_l,\ P(\bar{\phi}) \ge 0, \Pi(\bar{\phi}) \ge 0, \bar{\phi} \ge 0,$$

where ϕ_l is concentration of an i-th substance, $P_l(\bar{\phi})$ is destruction functional and $\Pi_l(\bar{\phi})$ is production functional. In [3] a family of unconditionally monotonic schemes have been built, from the first to fourth order of accuracy. One of the first order schemes is equivalent to the known QSSA scheme [4]

$$\phi_l^j = \phi_l^{j-1}e^{-P_l(\phi^{j-1})\tau} + \frac{1 - e^{-P_l(\phi^{j-1})\tau}}{P_l(\phi^{j-1})\tau}(\Pi_l(\phi^{j-1}) + r_l^j)\tau.$$

Data assimilation result is sought as the minimum of

$$\Phi(\bar{\phi}^j) = \sum_{l=1}^{Nc}(\phi_l^j - I_l^j)^2 M_l + \alpha\sum_{l=1}^{Nc}(r_l^j)^2,$$

where M_l is equal to 1 if l-th substance is measured and zero otherwise. This minimum is given by the formula:

$$\phi_l^j = \frac{1}{1+Z_l}\left(\phi_l^j e^{-P_l\left(\bar{\phi}^{j-1}\right)\tau} + \frac{1 - e^{-P_l\left(\bar{\phi}^{j-1}\right)\tau}}{P_l\left(\bar{\phi}^{j-1}\right)\tau}\left(\Pi_l\left(\bar{\phi}^{j-1}\right)\right)\tau\right) + \frac{Z_l}{1+Z_l}I_l^j,\ Z_l = \frac{M_l}{\alpha}\left(\frac{1 - e^{-P_l\left(\bar{\phi}^{j-1}\right)\tau}}{P_l\left(\bar{\phi}^j\right)\tau}\right)^2.$$

Splitting and data assimilation

Given system state on a previous time step ϕ^{j-1} Data assimilation problem solution is sought as a minimum of functional $J(\phi^j, r^j)$ on constraints presented by a model.

- Variational data assimilation to a generic model (without splitting):

$$J(\phi^j, r^j) = \left\|H\phi^j - I^j\right\|^2 + \alpha\left\|r^j\right\|^2, \quad \frac{\phi^j - \phi^{j-1}}{\tau} = L_x\phi^j + L_y\phi^j + r^j.$$

- Variational data assimilation to a split model as a whole (Total optimization):

$$J(\phi^j, r^j) = \left\|H\phi^j - I^j\right\|^2 + \alpha\left\|r^j\right\|^2, \quad \gamma_\xi\frac{\phi_\xi^j - \phi^{j-1}}{\tau} = L_\xi\phi_\xi^j + \gamma_\xi r^j, \quad \xi \in \{x, y\}, \quad \phi^j = \sum_\xi \gamma_\xi\phi_\xi^j.$$

  Stationarity conditions of the augmented functional:

$$\frac{\phi_\xi^j - \phi^{j-1}}{\tau} = L_\xi\phi_\xi^j + \gamma_\xi r^j, \quad 2\gamma_\xi H^*(H(\phi^j) - I^j) + \gamma_\xi\frac{\psi_\xi^j}{\tau} - L_\xi^*\psi_\xi^j = 0, \quad \xi \in \{x, y\},$$

$$\nabla_{r^j} J(\phi^j, r^j) = 2\alpha r^j - \sum_\xi \gamma_\xi\psi_\xi^j = 0.$$

  The system above can be solved iteratively:

$$r^j(0) = 0, \quad r^j(l+1) = r^j(l) - \beta_l\nabla_{r^j}J(\phi^j, r^j(l)).$$

- Fine-grained data assimilation [5–7]:

$$J(\phi_x^j, \phi_y^j, r_x^j, r_y^j) = \sum_\xi\left(\left\|H\phi_\xi^j - I^j\right\|^2 + \alpha\left\|r_\xi^j\right\|^2\right),$$

$$\gamma_\xi\frac{\phi_\xi^j - \phi^{j-1}}{\tau} = L_\xi\phi_x^j + r_\xi^j, \quad \xi \in \{x, y\}, \quad \phi^j = \sum_\xi\gamma_\xi\phi_\xi^j, \quad r^j = \sum_\xi r_\xi^j.$$

  Stationarity conditions of the augmented functional are independent for different dimensions:

$$\begin{cases} 2H^*(H\phi_\xi^j - I^j) + \gamma_\xi\frac{\psi_\xi^j}{\tau} - L_\xi^*\psi_\xi^j = 0 \\ 2\alpha r_\xi^j - \psi_\xi^j = 0, \\ \gamma_\xi\frac{\phi_\xi^j - \phi^{j-1}}{\tau} - L_\xi\phi^j - r_\xi^j = 0, \end{cases} \quad , \ \xi \in \{x, y\}$$

Fine-grained vs total optimization

Scenario description: Comparison of different DA schemes with splitting in 2D Domain (30km x 30km). Synthetic measurements are taken on a regular grid in space and time. **A priori sources are considered to be zero.**

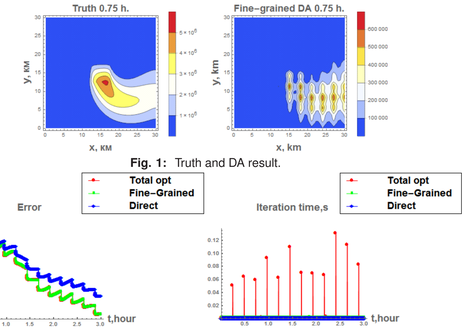

Fig. 2: DA algorithms comparison. Left: Error of the DA reult with respect to "truth" Right: Iteration time

Remark: Total optimization algorithm estimated both state and emission rate while fine-grained DA algorithm estimated the state with almost the same precision but its emission rate estimate was less accurate.

Vertical composition data assimilation scenario

Scenario description: Synthetic data case. Domain height is 3 km. Duration is 24 hours. In the middle of domain there is an additional NO source. NO_2 and O_3 are measured every 2 hours in every 400m height. Meteorological parameters are taken from real measurements.

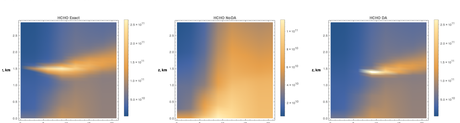

Fig. 3: Comparison of "truth", background forecast and DA solution for $HCHO$ that is not assimilated.

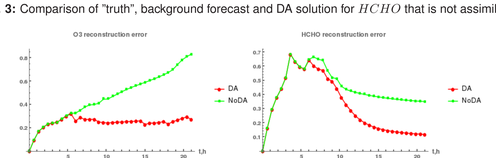

Fig. 4: DA error dynamics for assimilated substance and a substance that is not assimilated.

2D Chemical data assimilation scenario (22 substances)

Scenario description: The scenario has been prepared together with R.Nuterman and A.Mahura (DMI) within COST Action ES1004 [9]. It is based on [8]. Consider a domain $(X = 61{*}83km, Y = 61{*}81km, T = 3720{*}720s = 31day)$ for July 2010. Temporal grid for chemical kinetics is 100 times finer.
Parameters: Wind velocities has been taken from the Integrated EnviroHIRLAM model [12]. Initial conditions has been generated with MOZART-IFS model [13]. **A priori sources are considered to be zero.**
Measurement data: Airbase [14] data of 198897 SO_2, O_3, NO_2, CO, NO measurements are assimilated and 197854 data are used for reference.

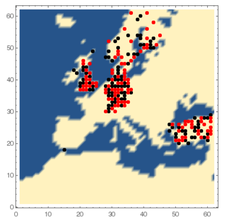

Fig. 4: Airbase measurement sites. Black are assimilated and red are used for validation

Evaluation procedure: Part of sites are assimilated and the other part is used for evaluation.

Subst	TrspTrns	$\sigma = 0$	$\sigma = 0.3$	$\sigma = 0.7$
O_3	0.710335	0.944806	0.943968	0.941273
NO_2	0.517428	0.825647	0.824212	0.819671
NO	0.443524	0.82297	0.822155	0.819529
CO	0.258675	0.658969	0.662524	0.66599
SO_2	0.22953	0.809995	0.8083	0.804114

Table. 1: Correlation for synthetic data with different noise levels

Subst	TrspTrns	DATrspTrns
O_3	-0.113619	0.489755
NO_2	0.0896602	0.211484
NO	0.0849778	0.220452
CO	0.0497666	0.017484
SO_2	-0.0104786	0.0733444

Table. 2: Correlation for real data

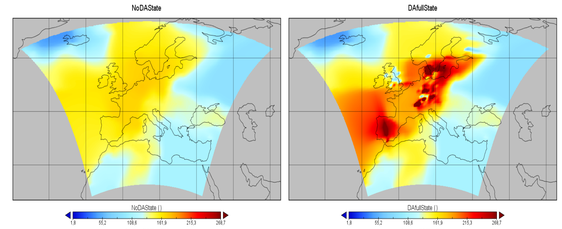

Fig. 5: Background forecast (Left), DA (Right).

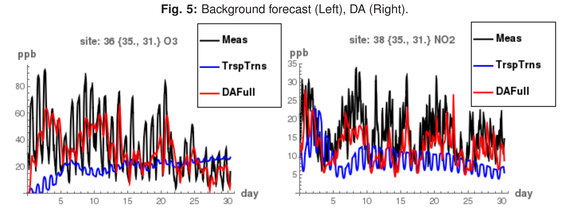

Fig. 6: Comparison of measurement data, data assimlation result and background forecast on a reference site.

Conclusions

- Combining splitting schemes and data assimilation schemes allows to construct computationally effective algorithms without iterations for data assimilation of in situ measurements to convection-diffusion-reaction models.
- In numerical experiments fine grained data assimilation scheme for split model showed almost the same precision as the conventional scheme being more computationally effective. Both schemes showed better performance than direct data inclusion method.
- Fine-grained scheme was less effective as a source reconstruction method.
- In the considered real measurement data case the DA algorithm showed better performance than direct model for O_3 concentrations and less improvement for NO.
- In a synthetic data case with the "unknown" NO source DA has improved the result significantly.

Acknowledgements

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.