View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Tomsk State University Repository

Direct variational assimilation algorithm for atmospheric chemistry data with transport and transformation model

Alexey Penenko[1,2],

Vladimir Penenko[1,2],

Roman Nuterman[3,4,5],

Alexander Baklanov[3,6],

Alexander Mahura[3]

[1] *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, ICM&MG SB RAS, prospekt Akademika Lavrentjeva 6, 630090, Novosibirsk, Russia*

[2] *Novosibirsk State University, NSU, Pirogova Str. 2, 630090, Novosibirsk, Russia*

[3] *Danish Meteorological Institute, DMI, Lyngbyvej 100, DK-2100, Copenhagen, Denmark*

[4] *University of Copenhagen, Nørregade 10, PO Box 2177, 1017 Copenhagen K, Denmark*

[5] *Tomsk State University, TSU, Lenin Ave., 36, 634050, Tomsk, Russia*

[6] *World Meteorological Organization, WMO, 7bis av. la Paix, CP2300, CH-1211, Geneva, Switzerland*

ABSTRACT

Atmospheric chemistry dynamics is studied with convection-diffusion-reaction model. The numerical Data Assimilation algorithm presented is based on the additive-averaged splitting schemes. It carries out "fine-grained" variational data assimilation on the separate splitting stages with respect to spatial dimensions and processes i.e. the same measurement data is assimilated to different parts of the split model. This design has efficient implementation due to the direct data assimilation algorithms of the transport process along coordinate lines. Results of numerical experiments with chemical data assimilation algorithm of *in situ* concentration measurements on real data scenario have been presented. In order to construct the scenario, meteorological data has been taken from EnviroHIRLAM model output, initial conditions from MOZART model output and measurements from Airbase database.

Chemical data assimilation, variational approach, advection-diffusion-reaction model, fine-grained data assimilation, splitting method, discrete-analytical schemes

1. INTRODUCTION

We consider the following classes of problems associated to the inverse modeling:

- Direct problems: System’s behavior has to be forecasted and studied with a mathematical model and prescribed parameters.

21st International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics,
edited by G. G. Matvienko, O. A. Romanovskii, Proc. of SPIE Vol. 9680, 968076
© 2015 SPIE · CCC code: 0277-786X/15/$18
· doi: 10.1117/12.2206008

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

- Inverse problems: Model parameters must be adjusted to fit model forecasts to the corresponding measurement data. It may take to solve series of direct problems with various model parameters.
- Data assimilation (DA) problems: A forecast has to be improved (on-line) by adjusting model parameters with incoming measurement data. It may take to solve series of inverse problems with various measurement data.

In the work, we present data assimilation algorithm for advection-diffusion-reaction atmospheric chemistry model. To construct a data assimilation algorithm, the following features should be taken into account:

- Atmospheric composition is being changed rapidly, therefore current and future system state is of interest.
- Stiff chemical kinetics equations (different time scales), various chemical mechanisms and their nonlinear behavior.
- Uncertainties are not only in initial conditions but also in model coefficients (reaction rates) and in emission rates.
- High dimensionality ($\approx 10^7$) of modern atmospheric chemistry transport models due to high number of spatial variables and different species, imposes requirements to the computational performance.
- Relatively small number of chemical species in a small number of spatial points can be measured.
- Data assimilation algorithms must be embedded in existing models.
- Multidisciplinary study.

A review and examples of chemical data assimilation algorithms can be found in[1,2,3]. Summarizing them, we would like to emphasize that unlike data assimilation in meteorology initial states in the chemical data assimilation are to be "forgotten" due to diffusion process. Meanwhile the emission rates and model coefficients play a significant role as the sources of uncertainty in the chemical data assimilation. In our work we use source-term uncertainty to perform data assimilation.

In order to apply a Data Assimilation algorithm to real data, one has to prepare a consistent set of parameters:

- Meteorological conditions for transport and transformation of model parameters.
- Chemical background for initial and boundary conditions.
- Measurement data have to be reduced to a common form and divided into assimilated and reference sets.

The data assimilation algorithm has to be applied to the assimilated set and compared to a reference one.

2. DATA ASSIMILATION ALGORITHM

2.1. Transport and transformation model

Let us consider a spatial-temporal domain:

$$\vec{x} = (x_1, x_2, x_3) \in \Omega = [0, l_1] \times [0, l_2] \times [0, l_3],\ t \in [0, T],\ \Omega_T := \Omega \times [0, T],$$

bounded by $\delta\Omega_T = \delta\Omega \times [0, T]$. In the domain we consider atmospheric chemistry transport and transformation model for different substances like contaminants, heat, moisture, radiation, etc.

$$L\bar{\phi} \equiv \frac{\partial \bar{\phi}(\vec{x}, t)}{\partial t} + \operatorname{div}\left(\vec{u}\,\bar{\phi}(\vec{x}, t) - \mu(\vec{x}, t)\operatorname{grad}\,\bar{\phi}(\vec{x}, t)\right) =$$

$$= S(\bar{\phi}(\vec{x}, t)) + \bar{f}(\vec{x}, t) + \bar{r}(\vec{x}, t),\ (\vec{x}, t) \in \Omega_T, \qquad (1)$$

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

$$\mu(\vec{x},t)\frac{\partial\bar{\phi}(\vec{x},t)}{\partial\vec{n}}+\beta(\vec{x},t)\bar{\phi}(\vec{x},t)=\bar{g}(\vec{x},t),\ (\vec{x},t)\in\partial\Omega_T, \quad (2)$$

$$\bar{\phi}(\vec{x},0)=\bar{\phi}^0(\vec{x}),\ \vec{x}\in\Omega. \quad (3)$$

Here $\bar{\phi}(\vec{x},t)$ is a state function that has physical meaning of concentrations fields at point $(\vec{x},t)\in\Omega_T$, e.g. $\phi_l(\vec{x},t)$ corresponds to the concentration of l^{th} substance at point $(\vec{x},t)$. Here $l=1,\dots,N_c$, N_c is the number of considered substances. Vector $\vec{u}(\vec{x},t)=(u_1(\vec{x},t),u_2(\vec{x},t),u_3(\vec{x},t))$ denotes "wind speed", $\mu(\vec{x},t)=diag(\mu_1(\vec{x},t),\mu_2(\vec{x},t),\mu_3(\vec{x},t))$ is a diagonal diffusion tensor, $S:\mathbb{R}^{N_c}\to\mathbb{R}^{N_c}$ is a transformation operator, $\vec{n}$ is the boundary outer normal direction, $\bar{f}(\vec{x},t)$, $\bar{g}(\vec{x},t)$, $\bar{\phi}^0(\vec{x})$ - a priori data for the sources and initial data, $\bar{r}(\vec{x},t)$ is a control function (uncertainty), that is introduced in the perfect model structure to assimilate data. As it is for $\bar{\phi}(\vec{x},t)$, each entry of $\bar{f}(\vec{x},t)$, $\bar{g}(\vec{x},t)$, $\bar{r}(\vec{x},t)$, $\bar{\phi}^0(\vec{x})$ vectors corresponds to a quantity attributed to l-th substance at point $(\vec{x},t)$.

Transformation operator S is defined by the chemical kinetics system of 22 reacting species from[4,5] augmented with the SO_2 reaction taken from the CMAQ model[6]:

$\mathrm{hv+NO2\to NO+O3P}$	$\mathrm{hv+O3\to O1D+O2}$
$\mathrm{HCHO+hv\to CO+2.HO2}$	$\mathrm{HCHO+hv\to CO+H2}$
$\mathrm{O2+O3P\to O3}$	$\mathrm{N2+O1D\to N2+O3P}$
$\mathrm{O1D+O2\to O2+O3P}$	$\mathrm{H2O+O1D\to 2.OH}$
$\mathrm{HO2+NO\to NO2+OH}$	$\mathrm{NO+O3\to NO2+O2}$
$\mathrm{NO+RO2\to HCHO+HO2+NO2}$	$\mathrm{CO+OH\to CO2+HO2}$
$\mathrm{HC+OH\to H2O+RO2}$	$\mathrm{HCHO+OH\to CO+H2O+HO2}$
$\mathrm{NO2+OH\to HNO3}$	$\mathrm{2.HO2\to H2O2+O2}$
$\mathrm{H2O+2.HO2\to H2O+H2O2+O2}$	$\mathrm{HO2+RO2\to O2+ROOH}$
$\mathrm{2.RO2\to Prod}$	$\mathrm{OH+SO2\to HO2+Sulf}$

Reaction rates have been taken from[5] and depend on time, i.e, photochemistry is considered.

This kinetics system can be presented in the production-destruction operator form:

$$S_l(\vec{\phi}(\vec{x},t))=-P_l(\vec{\phi}(\vec{x},t))\phi_l(\vec{x},t)+\Pi_l(\vec{\phi}(\vec{x},t)),l=1,\dots,N_c, \quad (4)$$

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

Direct problem: With given $\vec{f}$, $\vec{g}$, $\vec{\phi}^0$, $\vec{r}$ determine $\vec{\phi}$ from (1)-(3). Exact solution $\vec{\phi}_*$ is a solution of direct problem corresponding to "unknown" emissions $\vec{r}_*$.

We consider all the functions and model parameters to be smooth enough for the solutions to exist and further transformations to make sense.

For the numerical solution let us introduce uniform temporal grid $\omega_t=\{t^j\}_{j=1}^{N_t}$ on $[0,T]$ with step size τ and N_t points and uniform spatial grids ω_β with N_β, $\beta=1,2,3$ grid points on Ω, $\omega=\omega_1\times\omega_2\times\omega_3$. Let $Q(\omega)$ be the space of real grid functions on ω. Direct problem can be efficiently solved with splitting method. Let us consider additive-averaged splitting scheme (analogous to[8]) on the intervals $t^j\le t\le t^{j+1}$. The splitting is done with respect to physical process (advection-diffusion and transformation processes) and advection-diffusion part is further split with respect to spatial dimensions. Finally we have 4 parallel stages for the step partition $\sum_{\beta=1}^4\gamma_\beta=1$ and sources partition $\bar{f}=\sum_{\beta=1}^4\bar{f}_\beta$.

- Convection-diffusion processes ($\beta=1,2,3$)

$$\gamma_\beta\frac{\partial\vec{\phi}_\beta(\vec{x},t)}{\partial t}+\frac{\partial}{\partial x_\beta}\left(u_\beta(\vec{x},t)\vec{\phi}_\beta(\vec{x},t)\right)-\frac{\partial}{\partial x_\beta}\left(\mu_\beta(\vec{x},t)\frac{\partial}{\partial x_\beta}\vec{\phi}_\beta(\vec{x},t)\right)$$
$$=\overrightarrow{f_\beta}(\vec{x},t)+\vec{r}_\beta(\vec{x},t),\ (\vec{x},t)\in\Omega\times[t^{j-1},t^j],$$

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

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

where $\partial\Omega_\beta=\{\vec{x}\in\partial\Omega|\ x_\beta=0||x_\beta=l_\beta\}$. This initial value problem can be approximated with implicit matrix form:

$$\gamma_\beta\frac{\vec{\phi}_\beta^j-\vec{\phi}^{j-1}}{\tau}+L_\beta\vec{\phi}_\beta^j=\vec{r}_\beta^j+\vec{f}_\beta^j, \tag{6}$$

$$L_\beta(\vec{\phi}):=\{L_\beta(\vec{\phi}_l)\}_{l=1}^{N_c}. \tag{7}$$

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

- Chemical reaction processes ($\beta=4$)

$$\gamma_\beta\frac{\partial\vec{\phi}_\beta(\vec{x},t)}{\partial t}+diag(\vec{P}(\vec{\phi}_\beta(\vec{x},t)))\vec{\phi}_\beta(\vec{x},t)$$
$$=\vec{\Pi}(\vec{\phi}_\beta(\vec{x},t))+\vec{f_\beta}(\vec{x},t)+\vec{r}_\beta(\vec{x},t),$$
$$(\vec{x},t)\in\Omega\times[t^{j-1},t^j],$$
$$\vec{\phi}_\beta(\vec{x},t^j)=\vec{\phi}(\vec{x},t^j),\ \vec{x}\in\Omega,$$

or in the entry-wise form

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

where P_l is the destruction rate functional and Π_l is the production functional. In[8,9,10] a family of unconditionally monotonic schemes have been built, from the first to fourth order of accuracy. One of the single stage schemes is equivalent to the known QSSR scheme[11]:

$$\phi_{\beta l}^j(\bar{p})=A_l(\vec{\phi}^{j-1}(\bar{p}))+B_l(\vec{\phi}^{j-1}(\bar{p}))r_{\beta l}^j(\bar{p}),\ l=1,\dots,N_c,\ \bar{p}\in\omega,$$
$$A_l(\vec{\phi}^{j-1}(\bar{p}))=\phi_l^{j-1}(\bar{p})e^{-P_l(\vec{\phi}^{j-1}(\bar{p}))\tau}$$
$$+\frac{1-e^{-P_l(\vec{\phi}^{j-1}(\bar{p}))\tau}}{P_l(\vec{\phi}^{j-1}(\bar{p}))\tau}(\Pi_l(\vec{\phi}^{j-1}(\bar{p}))+f_{\beta l}^j(\bar{p}))\tau,$$
$$B_l(\vec{\phi}^{j-1}(\bar{p}))=\frac{1-e^{-P_l(\vec{\phi}^{j-1}(\bar{p}))\tau}}{P_l(\vec{\phi}^{j-1}(\bar{p}))\Delta t}\tau.$$

In vector form

$$\vec{\phi}_\beta^j=\vec{A}(\vec{\phi}^{j-1})+diag(\vec{B}(\vec{\phi}^{j-1}))\vec{r}_\beta^j. \tag{8}$$

- Next step approximation

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

$$\vec{\phi}^j = \sum_{\beta=1}^{4} \gamma_\beta \, \vec{\phi}_\beta. \tag{9}$$

Advantage of the scheme is that individual processes for each dimension are evaluated independently (in parallel).

2.2 Fine-grained data assimilation to the model

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

$$I(t) = H(t, \vec{\phi}^*(.,t)) + \vec{\eta}(t),\ t \in [0, T], \tag{10}$$

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

Data assimilation problem: Determine $\phi(.,t)$ for $t > t^*$ with (1)-(3), (10) and functionsf_a, g_a, ϕ_a^0, I defined on $0 < t \le t^*$.

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

$$\xi_m = \{(\bar{x}_M^m, t_M^m, l_M^m, I_m, \sigma_M^m)\},\ m = 1, \dots, N_M.$$

where $\bar{x}_M^m$ is the spatial coordinate of the measurement, t_M^m is the moment of measurement, l_M^m is the number of substance 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 in a time-step t^j we can use only measurements with $t_M^m \le t^j$. Let us define the set of indices

$$\theta^j = \{1 \le m \le N_M | t_M^m = t^j\}.$$

The corresponding measurement operator

$$H^j \vec{\phi} = \{\phi_{l_M^m}(\bar{x}_M^m, t_M^m)\}_{m \in \theta^j},$$
$$I^j = \{I_m\}_{m \in \theta^j},\ \sigma^j = \{\sigma_M^m\}_{m \in \theta^j}.$$

A function $\vec{\eta}(t)$ is from a set of admissible values that describe error estimate for measurement data. The error $\vec{\eta}$ is considered to be bounded in (weighted) Euclidean norm in the measurements space

$$\|\vec{\eta}(t)\|_{\sigma^j} = \sqrt{\sum_{m \in \theta^j} \left(\frac{\eta_m}{\sigma_M^m}\right)^2} \le \delta_{\vec{\eta}}.$$

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

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

where $\|.\|$ is the norm of a Hilbert space over $Q(\omega)^{N_c}$ and $\langle .,. \rangle$ is the corresponding inner product, α is the regularization (assimilation parameter), which selects whether the solution will be closer to the direct model solution or will reproduce measurements better. In the paper on the time step t^j we update only the control variable $\vec{r}^j$ for this time step. In the context of assimilating data to the split model we distinguish between the two approaches:

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

- In conventional approach[12,13] one takes minimum of $J^j(\vec{\phi}^j, \vec{r}^j)$ as the solution with constraints (6), (8), (9) and $\vec{r}_\beta = \gamma_\beta \vec{r}$, i.e., control functions of different splitting stages are connected.
- In the fine-grained approach[14,15,16] the same data are assimilated to different parts of model and the results are coupled afterwards. We seek for the minimum of the functional

$$J_f^j\left(\left\{\vec{\phi}_\beta^j, \vec{r}_\beta^j\right\}_{\beta=1}^4\right) = \sum_{\beta=1}^4 J^j\left(\vec{\phi}_\beta^j, \vec{r}_\beta^j\right)$$

on constraints (6), (8) with the independent $\vec{r}_\beta$.

Using the method of Lagrange multipliers to solve minimization problem with equality constraints, we can construct augmented functional:

$$\bar{J}_f(\left\{\vec{\phi}_\beta^j, \vec{r}_\beta^j\right\}_{\beta=1}^4) = \sum_{\beta=1}^4 J^j\left(\vec{\phi}_\beta^j, \vec{r}_\beta^j\right) + \sum_{\beta=1}^3 \left\langle \gamma_\beta \frac{\vec{\phi}_\beta^j - \vec{\phi}^{j-1}}{\tau} - L_\beta \vec{\phi}^j - \vec{r}_\beta^j - \vec{f}_\beta^j, \vec{\psi}_\beta^j \right\rangle + \left\langle \vec{\phi}_4^j - \vec{A}(\vec{\phi}^{j-1}) - diag\left(\vec{B}(\vec{\phi}^{j-1})\right)\vec{r}_4^j, \vec{\psi}_4^j \right\rangle$$

We can see that components of $\bar{J}_f(\left\{\vec{\phi}_\beta^j, \vec{r}_\beta^j\right\}_{\beta=1}^4)$ corresponding to different β are independent hence stationary point coordinates can be found independently.

In order to present an algorithm of finding a stable point for the convection-diffusion part, we need further elaboration of operator L. Because of splitting, we can consider equations (6) independent for each coordinate line in both dimensions. The algorithm is the same for any coordinate line and here we will describe the algorithm applied to a grid line along X axis for a fixed Y and Z indices $1 \le q \le N_2, 1 \le p \le N_3$ and l-th substance field. Let $\phi_i^j = \left(\left(\phi_\beta\right)_l\right)_{iqp}^j, f_i^j = \left(\left(f_\beta\right)_l\right)_{iqp}^j, r_i^j = \left(\left(r_\beta\right)_l\right)_{iqp}^j, 1 \le i \le N_\beta =: N$. For the sake of computational efficiency, we use approximations of (1) that produce tridiagonal matrix systems:

$$-a_i\phi_{i+1}^j + b_i\phi_i^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j, i = 0, \quad (11)$$

$$-a_i\phi_{i+1}^j + b_i\phi_i^j - c_i\phi_{i-1}^j = \phi_i^{j-1} + \tau r_i^j + \tau f_i^j,\ i = 1, \ldots, N-1, \quad (12)$$

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

In this term the assimilated state is the solution of the minimization problem

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

WRT (11)-(13) where M_i^j is the spatial-temporal measurement mask (i.e. M_i^j equals 1 if $\bar{x}_{iq} \in X_M^j$ and 0 otherwise, in other words, it is equal to 1 if there is a measurement data at point $\bar{x}_{iq}$), I_i^j is measurement data at point $\bar{x}_{iq}$ (if there is a measurement) and σ_i is measurement device standard deviation of the measurement in point $\bar{x}_{iq}$ (if there is a measurement). Introducing Lagrange multipliers, we obtain augmented functional:

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

Taking the first variations of the augmented functional equal to zero, we obtain the following equations:

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

$$\nabla_{\psi_i^j}\bar{J}_f(\phi^j, r^j, \psi^j) = 0,$$

which is equivalent to (11)-(13).

$$\nabla_{\phi_i^j}\bar{J}_f(\phi^j, r^j, \psi^j) = 0$$

is equivalent to

$$-c_{i+1}\psi_{i+1}^j + b_i\psi_i^j = -\frac{2M_i^j}{\sigma_i^2}\left(\phi_i^j - I_i^j\right)\tau,\ i = 0,$$
$$-c_{i+1}\psi_{i+1}^j + b_i\psi_i^j - a_{i-1}\psi_{i-1}^j = -\frac{2M_i^j}{\sigma_i^2}\left(\phi_i^j - I_i^j\right)\tau,\ i = 1, \dots, N-1,$$
$$b_i\psi_i^j - a_{i-1}\psi_{i-1}^j = -\frac{2M_i^j}{\sigma_i^2}\left(\phi_i^j - I_i^j\right)\tau,\ i = N.$$

and

$$\nabla_{r_i^j}\bar{J}_f(\phi^j, r^j, \psi^j) = 0$$

is equivalent to

$$2\alpha r_i^j - \psi_i^j = 0,\ i = 0, \dots, N.$$

The systems obtained can be merged into tridiagonal matrix equation[14,15,16]

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

which is solved by the direct matrix sweep method.

For the transformation step data assimilation the algorithm is the same for any grid point $\bar{p} \in \omega$. For brevity let $\vec{\phi}^j = \vec{\phi}_4^j(\bar{p}) \in \mathbb{R}^{N_c}$, $\vec{r}^j = \vec{r}_4^j(\bar{p}) \in \mathbb{R}^{N_c}$, $\vec{\psi}^j = \vec{\psi}_4^j(\bar{p}) \in \mathbb{R}^{N_c}$ and the result is sought as the stationary point of the augmented functional

$$\bar{J}(\vec{\phi}^j, \vec{r}^j) = \sum_{l=1}^{Nc}\left(\frac{\phi_l^j - I_l^j}{\sigma_l}\right)^2 M_l^j + \alpha\sum_{l=1}^{Nc}(r_l^j)^2 +$$
$$\sum_{l=1}^{Nc}\left(\phi_l^j - \phi_l^{j-1}e^{-P_l(\vec{\phi}^{j-1})\Delta t} - \frac{1 - e^{-P_l(\vec{\phi}^{j-1})\Delta t}}{P_l(\vec{\phi}^{j-1})\Delta t}(\Pi_l(\vec{\phi}^{j-1}) + r_l^j)\tau\right)\psi_l^j.$$

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

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

$$\phi_l^j = \frac{1}{1+Z} \left. \phi_l^{j-1} e^{-P_l(\vec{\phi}^{j-1})\tau} + \frac{1-e^{-P_l(\vec{\phi}^{j-1})\tau}}{P_l(\vec{\phi}^{j-1})\tau} \Pi_l(\vec{\phi}^{j-1})\tau \right) + \frac{Z}{1+Z} I_l^j,$$

$$Z = \frac{M_l^j}{\alpha(\sigma_l)^2} \left. \frac{1-e^{-P_l(\vec{\phi}^{j-1})\tau}}{P_l(\vec{\phi}^{j-1})\tau} \right)^2.$$

As we can see, the resulting algorithm can be implemented without iterations.

3. CHEMICAL DATA ASSIMILATION

3.1. Data assimilation scenario

In our numerical experiments we have done the following simplifications with respect to the data assimilation problem statement:

- The model used is a 2D model.
- Zero Neumann boundary conditions are used
- Simple transformation mechanism (22 chemical reactions).
- Transformation rates depend only on time of a day (important to account for pressure and temperature).
- Available emission databases are needed (annual update is desirable for operational use).
- A simplified diffusion coefficient is used.

In the data assimilation, we have used the domain, which is defined by temporal and spatial grids:

- Temporal grids:
  - Physical Time span: 1 July 2010 - 1 August 2010.
  - Temporal grid for transport processes
    $$Lt = 3720(grids) * 720(s) = 31day.$$
  - Nested temporal grid for transformation processes is 100 times finer:
    $$Lt = 372000(grids) * 7.2(s) = 31day.$$
- Spatial grids: (5 times coarser than Enviro-HIRLAM grid in each dimension):
  $$Lx = 61(grids) * 83(km) = 5086km,$$
  $$Ly = 61(grids) * 81(km) = 4962km.$$

Transport model parameters have been taken from the Enviro-HIRLAM[17] meteorological output. We have used wind speeds at 10m (U10,V10). Diffusion coefficient is evaluated from the wind speeds. Initial conditions have been taken from the MOZART[18] model output.

Measurement data have been taken from AirBase - the European Air quality database[19]. We have used 64001 SO_2,O_3,NO_2,CO,NO measurements in total from 22 Scandinavian (DK, SE, FI, NO) mea surement sites (Fig.1).

To study impact of both transformation and data assimilation processes, we have considered four configurations encoded by:

- DATrspT(rns): Transformations with data assimilation to transport processes.
- TrspTrns: Transformations with transport processes without data assimilation.
- DATrsp: Data assimilation to transport processes without transformation.
- Trsp: Transport processes (with neither transformation nor data assimilation).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

To compare configurations, some part of measurements are assimilated the rest is used as reference. The choice of division criteria defines a numerical experiment. Results are compared with respect to correlation coefficient and RMSE.

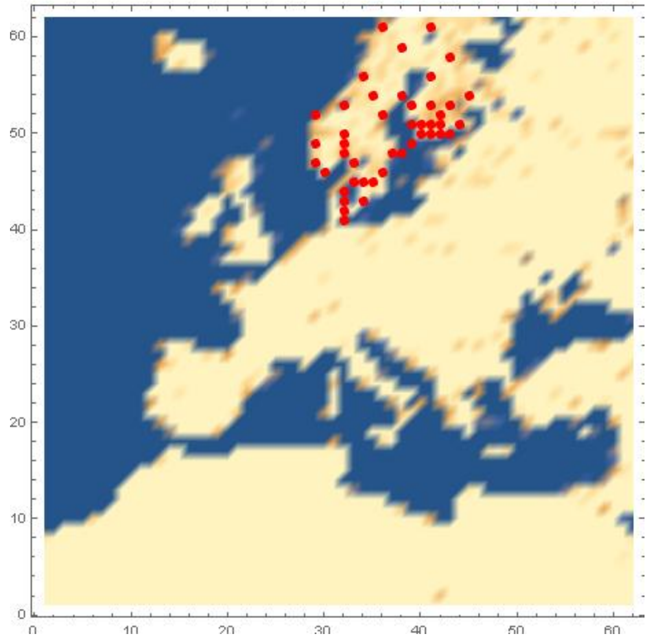

Figure 1. Computation domain and measurement sites locations (red dots).

3.2 Temporal difference

Data division criteria in the experiment: Assimilate data for t<T/2 and compare the model output to measurement data for t>T/2.

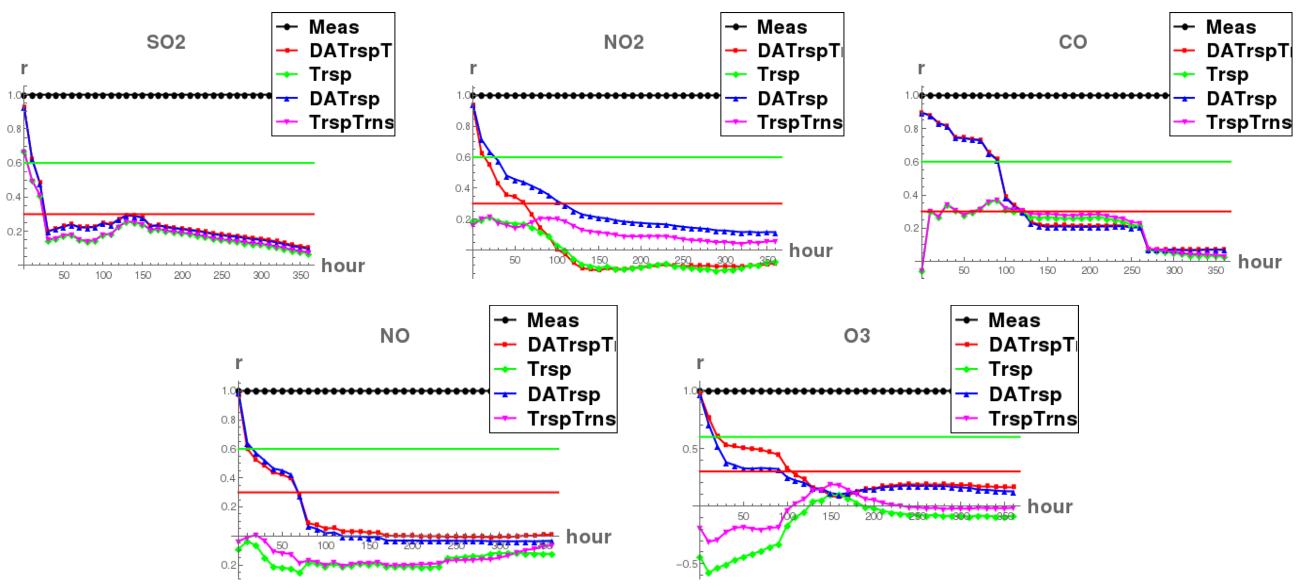

Figure 2. Correlation decrease with time after the last assimilated measurement.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

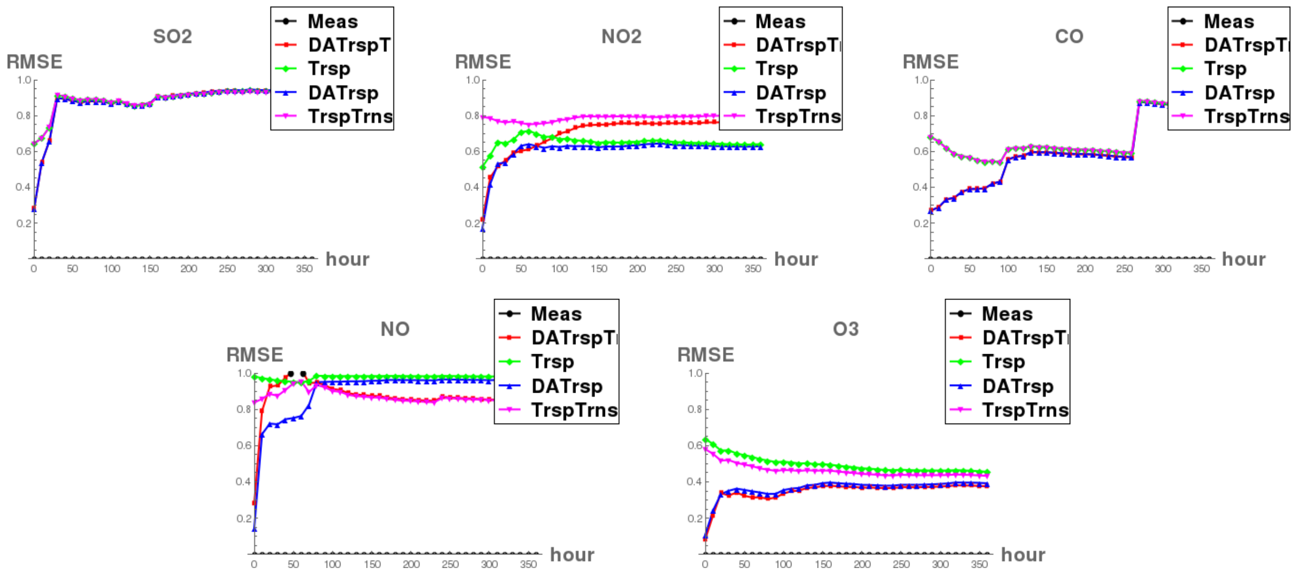

Figure 3. RMSE increase with time after the last assimilated measurement.

In Fig. 2 we presented results of comparison to time extending subsets of data from the reference set. From Fig. 2 we can conclude that the models with data assimilation (red and blue curves) provide better performance as the forecast tool compared to the direct models (Green and Magenta curves) for all the species except for SO_2. Based on this we can draw a general conclusion on the advantage of data assimilation in this case.

As we can see from the figures, the model with chemistry provides slower degradation of the correlation coefficient for O_3, while it is reverse for NO_2. For the rest of substances correlation degradation rate is almost identical.

Looking at RMSE degradation in Fig. 3 we can confirm the conclusion of Data assimilation benefits. Comparing DATrsp and DATrspT, we can state that RMSE degradation rate (its increase) for DATrspT for all species is higher than for DATrsp. Probably it is because of more uncertain and nonlinear nature of DATrspT with chemical transformations.

3.3 Species exclusion experiment

Data division criteria in the experiment: Assimilate all the data except for data of the selected substance.

Specie	Meas	DATrspTrns	Trsp	DATrsp	TrspTrns
CO	1.	0.0420364	0.0417183	0.0417183	0.0420361
NO	1.	0.370355	-0.068419	-0.068419	-0.0351235
NO_2	1.	0.458236	0.0633213	0.0633213	-0.019848
SO_2	1.	-0.0141014	-0.0160411	-0.0160411	-0.0141014
O_3	1.	0.119529	-0.058031	-0.058031	-0.0296002

Table 1. Comparison of results for different excluded species

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

As we can see from Table 1, the best performance for DA with chemical transformations has been obtained for NO_2 and NO. This can be explained by a strong link between the two species. In case the concentration of one of them is known, the other can be reconstructed. Less improvement has been obtained for O_3 (below significant correlation level of 0.3). For CO and SO_2 the results of all configurations have been almost the same.

4. CONCLUSION

Combination of splitting and data assimilation schemes let us construct computationally effective algorithms for data assimilation of *in situ* measurements to convection-diffusion models.

A complete data assimilation scenario has been compiled with meteorological data from Enviro-HIRLAM model, initial concentration data from MOZART model and *in situ* measurement data from Airbase.

We carried out series of numerical experiments in which we tested DA algorithms on different divisions of measurement data into assimilated and reference datasets. Data assimilation was able to improve modeling results with imperfect (approximate) models and model parameters. The advantage of DA algorithm that includes chemical transformations was identified for O_3 concentrations modeling. The experiments have shown a link between NO and NO_2 concentrations with respect to data assimilation (concentrations of one of these substances can be reconstructed by assimilated concentrations of the other).

Among the future steps to improve data assimilation results we can identify:

- Inclusion of more realistic boundary conditions is required.
- Additional tuning is essential for coefficients of chemical reactions.
- Quality control of chemical data measurements at stations is recommended for excluding of ”extreme” data.
- Revision of implementation procedure/steps for chemical model is required.
- Additional evaluation of monthly and seasonal variability is needed.

5. ACKNOWLEDGMENTS

Authors are thankful to DMI Colleagues for constructive comments and discussions. Authors appreciate Irina Gerasimova’s help with the text of the paper. The work has been partially supported by COST Action ES1004 STSM Grant #21654, RFBR grants NN 14-01-31482 mol_a and 14-01-00125,

6. REFERENCES

[1] Sandu, A., and Tianfeng, C., “Chemical data assimilation - an overview,” Atmosphere 2, 426–463 (2011).

[2] Baklanov, A. et. al., “Online coupled regional meteorology chemistry models in Europe: current status and prospects,” Atmos. Chem. Phys. 14, 317–398 (2014).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

[3] Elbern, H., Strunk, A., Schmidt, H., Talagrand, O., “Emission rate and chemical state estimation by 4-dimensional variational inversion,” Atmos. Chem. Phys. 7, 3749-3769 (2007).

[4] Stockwell, W., and Goliff, W., “Comment on Simulation of a reacting pollutant puff using an adaptive grid algorithm by R. K. Srivastava et al.,” J. Geophys. Res. 107, 4643-4650 (2002).

[5] Gery, M.W., Whitten, G.Z., Killus, J.P., and Dodge, M.C., “A photochemical kinetics mechanism for urban and regional scale computer modeling,” J. Geophys. Res., 94, 12952-12956 (1989).

[6] Byun, D. W., and Schere, K. L.,: “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 (2006).

[7] Samarskii, A.A., Vabishchevich, P.N., [Computational Heat Transfer] Wiley Chichester, 1-850 (1995).

[8] Penenko, V., Tsvetova, E., “Variational methods for construction of monotone approximations for atmospheric chemistry models,” J.Comp.Appl.Math 6(3), 210-220 (2013).

[9] Penenko, V., Tsvetova, E., “Discrete-analytical methods for the implementation of variational principles in environmental applications,” J.Comp.Appl.Math 226, 319-330 (2009).

[10] Penenko, V.V., Tsvetova, E.A., Penenko, A.V., “Variational approach and Euler's integrating factors for environmental studies,” Comput. Math. Appl. 67(12), 2240–2256 (2014).

[11] 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. 10, 971–994 (1978).

[12] Penenko, V.V. , [Methods of numerical modeling of atmospheric processes,] Hydrometeoizdat Leningrad, 1-352 (1981)

[13] Marchuk, G.I., Zalesny, V.B., “A numerical technique for geophysical data assimilation problems using Pontryagin’s principle and splitting-up method,” Russ. J. Numer. Anal. M. 8(4), 311–326 (1993).

[14] Penenko, A., “Some theoretical and applied aspects of sequential variational data assimilation,” Comp. Tech. 11(2), 35-40 (2006).

[15] Penenko, V.V., “Variational methods of data assimilation and inverse problems for studying the atmosphere, ocean, and environment,” Num. Anal. and Appl. 2(4), 341-351 (2009).

[16] Penenko, A.V., Penenko, V.V., “Direct data assimilation method for convection-diffusion models based on splitting scheme,” Comp. Tech. 19(4), 69-83 (2014).

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

[18] Flemming, J., Inness, A., Flentje, H., Huijne,n V., Moinat, P., Schultz, M. G., and Stein, O.,: “Coupling global chemistry transport models to ECMWF’s integrated forecast system,” Geosci. Model Dev. 2, 253–265 (2009).

[19] Airbase, "European air quality database," http://www.eea.europa.eu/data-and-maps (30 July 2015).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 11/26/2015 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx