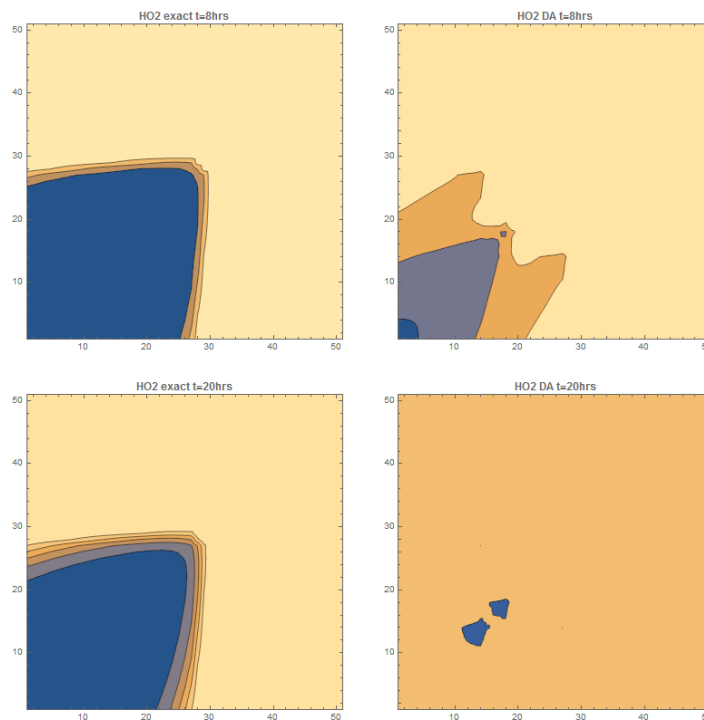# DMI
Ministry of Climate and Energy

Scientific Report 14-02

Discrete-Analytical Algorithms for Atmospheric Transport and Chemistry Simulation and Chemical Data Assimilation

Alexey Penenko [1], Vladimir Penenko [1],

Roman Nuterman [2,3,4], Alexander Mahura [2]

[1] *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, ICM&MG SB RAS, prospekt Akademika Lavrentjeva 6, 630090, Novosibirsk, Russia*
[2] *Danish Meteorological Institute, DMI, Lyngbyvej 100, DK-2100, Copenhagen, Denmark*
[3] *University of Copenhagen, Nørregade 10, PO Box 2177 , 1017 Copenhagen K, Denmark*
[4] *Tomsk State University, TSU, Lenin Ave., 36, 634050, Tomsk, Russia*

Copenhagen 2014

Colophon

Serial title:
Scientific Report **14-02**

Title:
Discrete-analytical algorithms for atmospheric transport and chemistry simulation and chemical data assimilation

Subtitle:

Author(s):
Alexey Penenko, Vladimir Penenko, Roman Nuterman, Alexander Mahura

Other contributors:

Responsible institution:
Danish Meteorological Institute

Language:
English

Keywords:
discrete-analytical schemes, splitting method, fine-grained data assimilation, chemical data assimilation, variational approach, convection-diffusion-reaction model

Url:
www.dmi.dk/dmi/sr14-02.pdf

Digital ISBN:
978-87-7478-654-2 (on-line)

ISSN:
1399-1949 (on-line)

Version:

Website:
www.dmi.dk

Copyright:
Danish Meteorological Institute
Application and publication of data and text is allowed with proper reference and acknowledgment

Content:

Abstract

Atmospheric chemistry dynamics is studied with convection-diffusion-reaction model. Measurement data is provided as the pointwise concentration measurements as it can be done by an automated city monitoring network. The numerical algorithm presented is based on the additive-averaged splitting scheme. It carries out "fine-grained" variational data assimilation on the convection-diffusion step while chemical reaction model is included as the forward model. This design provides efficient implementation due to the data assimilation scheme applied to the split transport process on each coordinate line containing contact measurement points. The goal of the work is to investigate the performance of the algorithm.

1. Introduction

The main purpose of the work was to implement data assimilation algorithm for a model problem of chemical weather forecast. Atmospheric chemistry processes can be modeled with convection-diffusion-reaction systems of partial differential equations. Correct simulation of reaction processes are of special interest because of nonlinearity and stiffness of corresponding differential equation systems. We investigate performance of discrete-analytical solver in atmospheric chemistry applications. It utilizes a special structure of equations to produce positive solutions with isolated exponential part. These solvers for chemistry are than used in frames of splitting technique with convection-diffusion part altered by embedded data assimilation algorithm. The main tasks were to compare (and cross-validate) algorithms for atmospheric chemistry simulation developed in ICM MG SB RAS (discrete-analytical) with the algorithms implemented within Enviro-HIRLAM model.

- First task were to investigate the system of equations modeling chemical kinetics implemented within Enviro-HIRLAM model. The scale of systems used in data assimilation should correspond to the amount of measurement data available (i.e. the system should be compact and representative). Within this task a system consisting of 20 chemical reactions (CBM-IV mechanism) and the Enviro-HIRLAM solver for its simulation have been studied.

- Second task was to implement a dicsrete-analytical solver for the chemical model determined on the first step. Solution produced by the solver has been compared to the one obtained with Enviro-HIRLAM solver with constant rates and rates changing during day-time.

- Third task was to add 2D atmospheric transport processes model to the chemical solver with splitting technique approach.

- Final task was to implement variational data assimilation algorithm for the constructed CBM-IV based convection-diffusion-reaction model.

Chemical data assimilation case has been constructed for CBM-IV mechanism. A direct convection-diffusion-reaction solver has been implemented. It is based on discrete-analytical scheme for stiff equations and has a property of preserving positive sign of the concentrations. Photochemical cycle has been taken into account by variable reaction rates. A solution matches the one obtained with Enviro-HIRLAM (Rosenbrock-based) solver. Variational data assimilation algorithm for 2D convection-diffusion-reaction model has been implemented and tested numerically with simulated data.

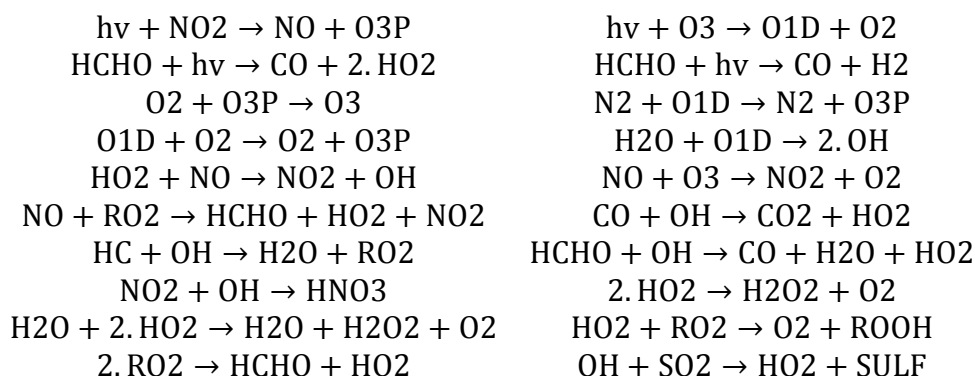# 2. Discrete-analytical solver for chemical kinetics system

2.1. Background

The following properties should be taken into account when modeling atmospheric chemistry:
- Chemical kinetics equations can be stiff (i.e. contain processes of wide range of temporal scales).
- Various chemical mechanisms should be considered.
- Physical and chemical sense of the solutions (e.g. positive concentrations) should be taken into account.
- Reaction rates can be spatially and temporarily non-homogeneous .
- High dimensionality ($\approx 10^7$) of modern models due to high number of spatial variables and different species, impose requirements to the computational performance.

2.2. Validation scenario

As the main example we have taken the following C(arbon)B(ound)M(echanism) from [2],[9],[5]. This solver-validation scenario include the following:

$$hv + NO2 \rightarrow NO + O3P \qquad hv + O3 \rightarrow O1D + O2$$
$$HCHO + hv \rightarrow CO + 2.\,HO2 \qquad HCHO + hv \rightarrow CO + H2$$
$$O2 + O3P \rightarrow O3 \qquad N2 + O1D \rightarrow N2 + O3P$$
$$O1D + O2 \rightarrow O2 + O3P \qquad H2O + O1D \rightarrow 2.\,OH$$
$$HO2 + NO \rightarrow NO2 + OH \qquad NO + O3 \rightarrow NO2 + O2$$
$$NO + RO2 \rightarrow HCHO + HO2 + NO2 \qquad CO + OH \rightarrow CO2 + HO2$$
$$HC + OH \rightarrow H2O + RO2 \qquad HCHO + OH \rightarrow CO + H2O + HO2$$
$$NO2 + OH \rightarrow HNO3 \qquad 2.\,HO2 \rightarrow H2O2 + O2$$
$$H2O + 2.\,HO2 \rightarrow H2O + H2O2 + O2 \qquad HO2 + RO2 \rightarrow O2 + ROOH$$
$$2.\,RO2 \rightarrow HCHO + HO2 \qquad OH + SO2 \rightarrow HO2 + SULF$$

This chemical model can be expressed in terms of ODE system:

$$CO'(t) = -CO(t)HO(t)k[[12]] + HCHO(t)HO(t)k[[14]] + \\ +HCHO(t)k[[3]] + HCHO(t)k[[4]], \tag{1}$$

$$CO2'(t) = CO(t)HO(t)k[[12]], \tag{2}$$

$$H2'(t) = HCHO(t)k[[4]], \tag{3}$$

$$H2O'(t) = -H2O(t)O1D(t)k[[8]] + HC(t)HO(t)k[[13]] + \\ +HCHO(t)HO(t)k[[14]], \tag{4}$$

$$H2O2'(t) = H2O(t)HO2(t)^{2\cdot}k[[17]] + HO2(t)^{2\cdot}k[[16]], \tag{5}$$

$$HC'(t) = -HC(t)HO(t)k[[13]], \tag{6}$$

$$HCHO'(t) = -HCHO(t)HO(t)k[[14]] - HCHO(t)k[[3]] - HCHO(t)k[[4]] + \\ +NO(t)RO2(t)k[[11]] + RO2(t)^{2\cdot}k[[19]], \tag{7}$$

$$HNO3'(t) = HO(t)NO2(t)k[[15]], \tag{8}$$

$$HO'(t) = -CO(t)HO(t)k[[12]] + 2.\,H2O(t)O1D(t)k[[8]] - HC(t)HO(t)k[[13]] - \\ -HCHO(t)HO(t)k[[14]] - HO(t)NO2(t)k[[15]] - HO(t)SO2(t)k[[20]] + \\ +HO2(t)NO(t)k[[9]], \tag{9}$$

$$HO2'(t) = CO(t)HO(t)k[[12]] - 2.\,H2O(t)HO2(t)^{2\cdot}k[[17]] + \\ +HCHO(t)HO(t)k[[14]] + 2.\,HCHO(t)k[[3]] + \\ +HO(t)SO2(t)k[[20]] - HO2(t)NO(t)k[[9]] - HO2(t)RO2(t)k[[18]] - \\ -2.\,HO2(t)^{2\cdot}k[[16]] + NO(t)RO2(t)k[[11]] + RO2(t)^{2\cdot}k[[19]], \tag{10}$$

$$N2'(t) = 0, \tag{11}$$

$$NO'(t) = -HO2(t)NO(t)k[[9]] - NO(t)O3(t)k[[10]] -$$
$$-NO(t)RO2(t)k[[11]] + NO2(t)k[[1]], \tag{12}$$

$$NO2'(t) = -HO(t)NO2(t)k[[15]] + HO2(t)NO(t)k[[9]] +$$
$$+NO(t)O3(t)k[[10]] +$$
$$+NO(t)RO2(t)k[[11]] - NO2(t)k[[1]], \tag{13}$$

$$O1D'(t) = -H2O(t)O1D(t)k[[8]] - N2(t)O1D(t)k[[6]] -$$
$$-O1D(t)O2(t)k[[7]] + O3(t)k[[2]], \tag{14}$$

$$O2'(t) = H2O(t)HO2(t)^2 \cdot k[[17]] + HO2(t)RO2(t)k[[18]] + HO2(t)^2 \cdot k[[16]] +$$
$$+NO(t)O3(t)k[[10]] - O2(t)O3P(t)k[[5]] + O3(t)k[[2]], \tag{15}$$

$$O3'(t) = -NO(t)O3(t)k[[10]] + O2(t)O3P(t)k[[5]]$$
$$- O3(t)k[[2]], \tag{16}$$

$$O3P'(t) = N2(t)O1D(t)k[[6]] + NO2(t)k[[1]] +$$
$$+O1D(t)O2(t)k[[7]] - O2(t)O3P(t)k[[5]], \tag{17}$$

$$RO2'(t) = HC(t)HO(t)k[[13]] - HO2(t)RO2(t)k[[18]] -$$
$$-NO(t)RO2(t)k[[11]] - -2.\,RO2(t)^2 \cdot k[[19]], \tag{18}$$

$$ROOH'(t) = HO2(t)RO2(t)k[[18]], \tag{19}$$

$$SO2'(t) = -HO(t)SO2(t)k[[20]], \tag{20}$$

$$SULF'(t) = HO(t)SO2(t)k[[20]] \tag{21}$$

Here time-depended variables denote concentrations of the corresponding species. Concentrations of $H2O$, $O2$ and $N2$ are supposed to be constant. Reaction rates corresponding to photochemical reactions (1)-(4) depend on solar radiation (Fig. 1).
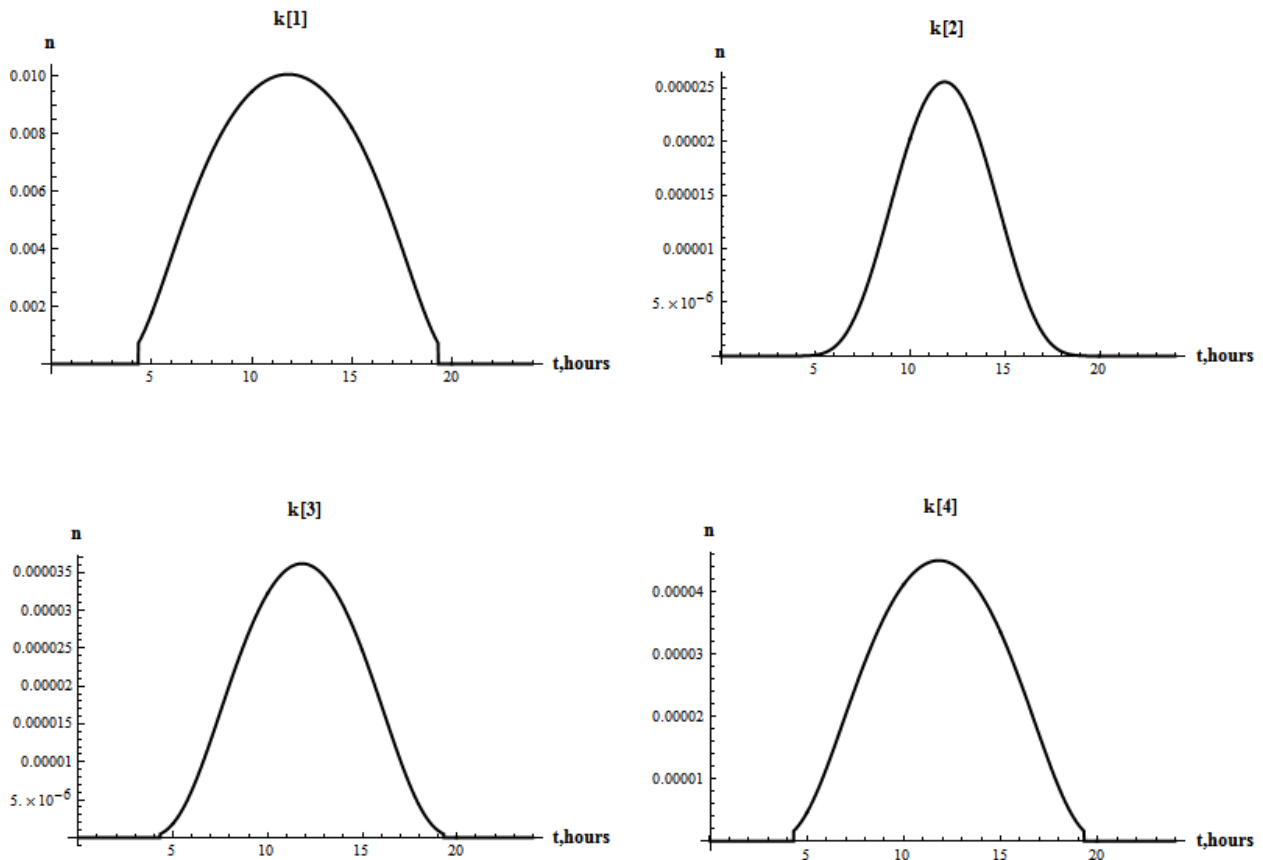

Figure 1. *Rates of photochemical reactions.*

The rest of the rates has been set to constants (see Table 1).

k(5)=1.60e-14	k(6)=2.60e-11
k(7)=4.03e-11	k(8)=2.20e-10
k(9)=7.91e-12	k(10)=1.55e-14
k(11)=3.99e-12	k(12)=2.19e-13
k(13)=2.13e-12	k(14)=9.21e-12
k(15)=9.75e-12	k(16)=3.23e-12
k(17)=9.37e-30	k(18)=1.51e-11
k(19)=3.31e-9	k(20)=9.27e-13

Table 1. Constant reaction rates.

2.3. Discrete-analytical scheme for chemical kinetics modeling

A review on the known methods for air chemistry modeling can be found at [15]. Here we use the way of scheme derivation from [8]. Universal ODE solvers can produce both positive and negative concentrations. Chemical sense can be preserved by taking into account the structure of the equation.

Consider an example

$$\text{HCHO}'(t) + \text{HCHO}(t)(k_{14}\text{OH}(t) + k_3 + k_4) =$$
$$= k_{11}\text{NO}(t)\text{RO2}(t) + k_{19}\text{RO2}(t)^{2.} \tag{22}$$

Equations have the following structure

$$\frac{\partial \phi}{\partial t} + P(\phi)\phi = \Pi(\phi), \quad P(\phi) \geq 0, \Pi(\phi) \geq 0, \phi \geq 0. \tag{23}$$

where ϕ is concentration of a substance, $P(\phi)$ is a destruction functional and $\Pi(\phi)$ is a production functional. Multiplying the equation by adjoint ϕ^* and integrating by parts on $t_j \leq t \leq t_{j+1}$, we obtain an identity, that holds for all ϕ^*

$$\phi(t_{j+1})\phi^*(t_{j+1}) - \phi(t_j)\phi^*(t_j) + \int_{t_j}^{t_{j+1}} \phi \left[-\frac{\partial \phi^*}{\partial t} + P(\phi)\phi^* \right] dt =$$
$$= \int_{t_j}^{t_{j+1}} \Pi(\phi)\phi^* dt. \tag{24}$$

If ϕ^* is taken as the solution of

$$-\frac{\partial \phi^*}{\partial t} + P(\phi)\phi^* = 0, \quad \phi^*(t_{j+1}) = 1, \tag{25}$$

then

$$\phi^{j+1} = \phi^j \phi^{*j} + \int_{t_j}^{t_{j+1}} \Pi(\phi)\phi^* dt.$$

If ϕ^* is approximated with solution of (25) for a constant coefficient $P(\phi^j)$, then

$$\phi^{j+1} = \phi^j e^{-P(\phi^j)\Delta t} + \int_{t_j}^{t_{j+1}} \Pi(\phi)e^{-P(\phi^j)(\Delta t - t)} dt$$

After the approximation of the integral term we obtain

$$\phi^{j+1} = \phi^j e^{-P(\phi^j)\Delta t} + \frac{1 - e^{-P(\phi^j)\Delta t}}{P(\phi^j)\Delta t} \Pi(\phi^j)\Delta t.$$

In this simple case of approximation the scheme is equivalent to the scheme known under QSSA name [3]. In [8] a family of unconditionally monotonic schemes have been built, from the first to fourth order of accuracy.

Automatic solver code generator has been developed on the Wolfram Research Mathematica Computer Algebra Platform and has features that are analogous to KPP Solver generation software.

It works in the following way
- Kinetic equations in textual "chemical" form are taken as the input.
- Automatic translation to the ODE Form is carried out with Wolfram Research Mathematica xCellerator package [10].
- ODE are parsed to (23) structure.
- Discrete-analytical scheme is generated for the structure.
- Numerical schemes in symbolic form are translated to C++ solver code.

2.4. Discrete-analytical solver validation

The solver created has been validated against the Rosenbrock-based implicit solver. On Fig. 2 magenta line corresponds to a KPP generated Rosenbrock solver and color lines correspond to the different number of inner iterations of discrete-analytical solver. Inner iterations are organized as a subgrid for original uniform grid with number of grid points $Nt = 1200$ on the time interval of 5 days.
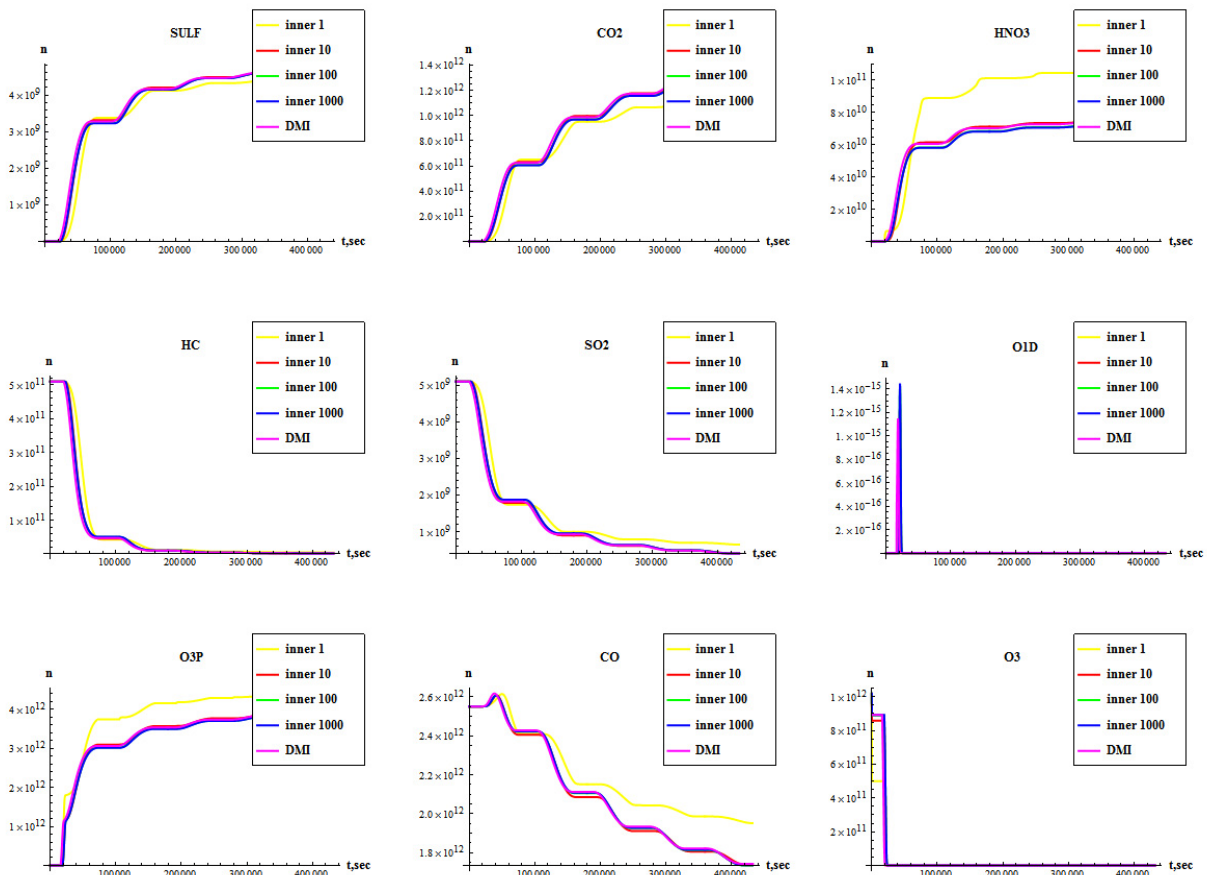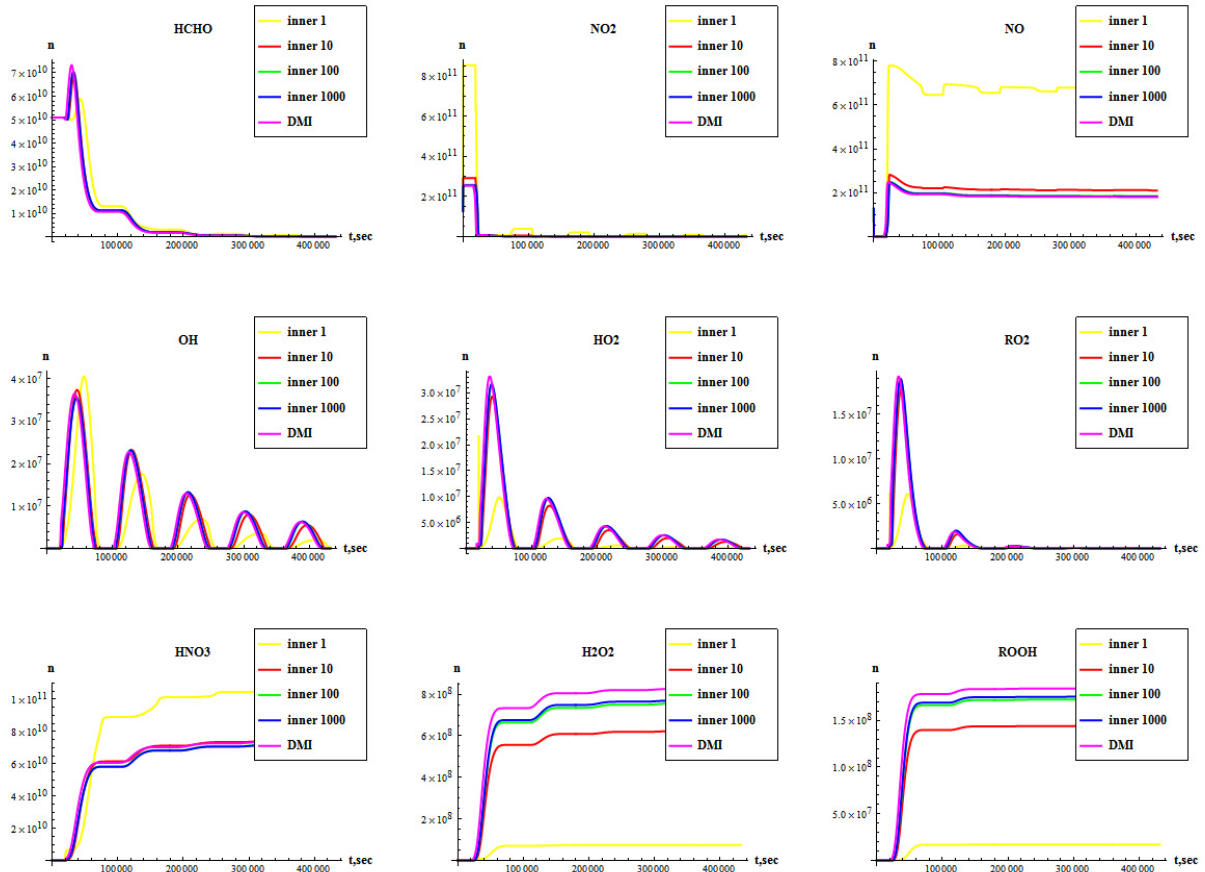
Figure 2. *Solution comparison for different number of inner iterations (from yellow to blue). Magenta denote KPP generated Rosenbrock solver. Time interval is 5 days.*

Analyzing the figures we can conclude that convergence for substances on Fig. 2 are satisfactory.

3. Chemical Data Assimilation

3.1. Data Assimilation Problems

Data assimilation algorithms aim to improve a forecast with the joint use of mathematical model and available measurement data. To construct the algorithm the following properties should be taken into account:

- Usually a measurement operator inversion result is non unique (Measurement data is not enough to find all the unknowns).
- Current (and future) state of the system is of interest.
- Solution should be obtained in real time as the situation becomes outdated very fast.
- Data assimilation algorithms should be embedded into the existing models.

A review on chemical data assimilation algorithms can be found at [11,1]. Summarizing it we would like to emphasize that unlike data assimilation in meteorology in chemical data assimilation initial states are to be "forgotten" due to diffusion process. Meanwhile the emission rates and model coefficients began to play a significant role as the sources of uncertainty in chemical data assimilation. In our work we use source-term uncertainty to carry out data assimilation.

3.2. Data assimilation problem statement

Consider a rectangular spatia-temporal domain:
$$\vec{x} = (x_1, x_2, x_3) \in \Omega = [0, l_1] \times [0, l_2] \times [0, l_3], \ t \in [0, T], \ \Omega_T := \Omega \times [0, T] \in \mathbb{R}^4$$
bounded by $\delta\Omega_T = \delta\Omega \times [0, T]$. In the domain we consider convection-diffusion-reaction model. It describes atmospheric transport and transformation processes for different substances (contaminants, heat, moisture, radiation etc.):

$$L\phi \equiv \frac{\partial \phi(\vec{x}, t)}{\partial t} + \text{div}\, (\vec{u}\phi(\vec{x}, t) - \mu(\vec{x}, t)\, \text{grad}\, \phi(\vec{x}, t)) =$$
$$= S(\phi(\vec{x}, t)) + f_a(\vec{x}, t) + r(\vec{x}, t),$$
$$(\vec{x}, t) \in \Omega_T, \tag{26}$$

$$\mu(\vec{x}, t)\frac{\partial \phi(\vec{x}, t)}{\partial \vec{n}} + \beta(\vec{x}, t)\phi(\vec{x}, t) = g_a(\vec{x}, t), \ (\vec{x}, t) \in \partial\Omega_T, \tag{27}$$

$$\phi(\vec{x}, 0) = \phi_a^0(\vec{x}), \ \vec{x} \in \Omega. \tag{28}$$

where $\phi(\vec{x}, t)$ is a statefunction, $\vec{u}(\vec{x}, t) = (u_1(\vec{x}, t), u_2(\vec{x}, t), u_3(\vec{x}, t))$ are "wind velocity" vector, $\mu(\vec{x}, t) = diag(\mu_1(\vec{x}, t), \mu_2(\vec{x}, t), \mu_3(\vec{x}, t))$ is a diagonal diffusion tensor, S is a transformation operator, $\vec{n}$ is the boundary outer normal direction, $f_a(\vec{x}, t), g_a(\vec{x}, t), \phi_a^0(\vec{x})$ - a priori data for the sources and initial data, $r(\vec{x}, t)$ is a control function (uncertainty), that is introduced in the rigid model structure to assimilate data.

Direct problem is to determine ϕ from (26)-(28) with known f_a, g_a, ϕ_a^0, r. Exact solution $\bar{\phi}$ is a solution of direct problem corresponding to an "unknown" set $\bar{f}, \bar{g}, \bar{\phi}^0, \bar{r}$.
Let us define measurement operator $\boldsymbol{H}$, that connects model state function with measurement data
$$\overrightarrow{\Psi}(t) = \boldsymbol{H}(\bar{\phi}(., t)) + \vec{\eta}(t), \ t \in [0, T], \tag{29}$$
where $\overrightarrow{\Psi}(t) \in \mathbb{R}^M$ are measurement data, M number of measurements in a time moment, $\vec{\eta} \in \mathbb{R}^M$ is a function from a set of admissible values that describes error estimate for measurement data. Error $\vec{\eta}$ is considered to be bounded in (weighted) Euclidean norm in measurements space $\|\vec{\eta}\|_M \leq \delta_{\vec{\eta}}$. We consider all the functions and model parameters to be smooth enough for the solutions to exist and transformations to make sense. Problem of $\phi(., t)$ determination for $t > t^*$ with (26)-(28) and (29) and functions $f_a, g_a, \phi_a^0, \overrightarrow{\Psi}$ defined on $0 < t \leq t^*$ is called a data assimilation problem.

3.3. Data assimilation scheme comparison

Usually data assimilation is carried out with "Explicit" scheme. In this case a direct problem is solved to obtain a "background forecast"

$$L\tilde{\phi}^{j+1} = \phi^j + \tau f^{j+1}, \quad \phi^{j+1} = \tilde{\phi}^{j+1} + K(H\tilde{\phi}^{j+1} - \Psi), \tag{30}$$

that is updated with measurement data

$$L\tilde{\phi}^{j+1} = \phi^j + \tau f^{j+1}, \quad \phi^{j+1} = \tilde{\phi}^{j+1} + K(H\tilde{\phi}^{j+1} - \Psi), \tag{31}$$

The solution of the problem can be presented as the minimum of Tikhonov functional:

$$J(\phi^{j+1}) = \|H(\phi^{j+1} - \tilde{\phi}^{j+1}) + H\tilde{\phi}^{j+1} - \Psi\|^2 +$$
$$+\alpha < (\phi^{j+1} - \tilde{\phi}^{j+1}), \phi^{j+1} - \tilde{\phi}^{j+1} >_K,$$

where $<.,.>_K$ is a weighted Cartesian product. The term $< (\phi^{j+1} - \tilde{\phi}^{j+1}), \phi^{j+1} - \tilde{\phi}^{j+1} >_K$ acts as a regularizer to the ill-posed measurement data inversion problem. In this case

- It is enough to solve direct problem (with existing codes).
- Regularizer norm has to be chosen wrt spatial solution structure (e.g. with ensemble methods, with many direct problem solutions).
- The structure is described with covariation matrices which indicate connected variables

The approach to Data assimilation that we use in the work can be called "Implicit". In this case measurement data is used to modify a control function that is introduced in the uncertain part of the model. Consider the following constrained optimization problem

$$J(\phi^{j+1}, r^{j+1}) = \|H\phi^{j+1} - \Psi\|^2 + \alpha\|r^{j+1}\|^2.$$

w.r.t. constraint linear model approximation

$$L\phi^{j+1} = \phi^j + \tau f^{j+1} + r^{j+1}.$$

Introducing Lagrange multipliers ϕ^{j+1^*}, a stationary point of the extended functional we can presented in form

$$\begin{pmatrix} L & -\dfrac{1}{2\alpha} \\ 2H^*H & L^* \end{pmatrix} \begin{pmatrix} \phi^{j+1} \\ \phi^{j+1^*} \end{pmatrix} = \begin{pmatrix} \phi^j + \tau f^{j+1} \\ 2H^*\Psi \end{pmatrix}. \tag{32}$$

In this case

- Direct problem (convection-diffusion) operator in the regularizer that diminishes "shock" effects of data inclusion (smoothes the solution).
- On the other hand one has to solve direct and adjoint problem system. In general case it takes iterations.

Solution ϕ^{j+1} of the system (32) also solves

$$(H^*H + \alpha L^*L)\phi^{j+1} = H^*\Psi + \alpha L^*(\phi^j + \tau f^{j+1})$$

and a minimum of the target functional

$$J(\phi^{j+1}) = \|H\phi^{j+1} - \Psi\|^2 + \alpha\|L\phi^{j+1} - (\phi^j + \tau f^{j+1})\|^2.$$

Assimilation parameter α regulates the ratio of the model and data in the solution [14].

- If the "weight" of the model increases than the solution converge to the direct solution and solution stability (wrt measurement noise) increases.
- If the model "weight" decreases then the data (and noise in it) is reproduced better.

Assimilation parameter can be chosen with Morozov discrepancy principle

$$\|H\phi^{j+1} - \Psi^{j+1}\| = \delta_*.$$

where δ_* is the measurement noise estimate. This estimate can be obtained from the statistical considerations. If measurement data is exact solution in a set of points perturbed by a Gaussian noise.

$$\Psi_m^j = \phi_{i(m)}^j + \sigma_m \xi_m, \ \xi_m \sim N(0,1),$$

then with probability p:

$$\delta_* = \sqrt{\sum_{m=1}^{M} \left(\frac{\Psi_m^j - \phi_{i(m)}^j}{\sigma_m} \right)^2} < \sqrt{\chi_{inv}^2(M,p)},$$

where $\chi_{inv}^2(M,p)$ - is the solution of

$$P(\chi_N^2 < \chi_{inv}^2(M,p)) = p.$$

I.e. data assimilation parameter is the probability "obtain" an upper estimate of the error. The bigger the probability the less the solution is affected by the measurements.

3.4. Splitting scheme

To solve multi-dimensional direct problem we use splitting method [4]. In order to use parallel computation technologies we will use additive-averaged splitting schemes (analogous to [12],[13]). Introducing temporal grid:

$$\omega_t = \{0 < \ldots < t_j < \ldots < t_{N_t} = T\}.$$

On each time interval the common convection-diffusion-reaction model (26)-(28) on subdomains $\Omega \times [t_{j-1}, t_j], j \geq 1$ can be presented as a splitting scheme, generated by decomposition

$$\sum_{k=1}^{3} \gamma_k + \gamma_c = 1 \tag{33}$$

$$\gamma_k \frac{\partial \phi_k(\vec{x},t)}{\partial t} + \frac{\partial (u_k(\vec{x},t)\phi_k(\vec{x},t))}{\partial x_k} - \frac{\partial}{\partial x_k}(\mu_k(\vec{x},t)\frac{\partial \phi_k(\vec{x},t)}{\partial x_k}) \tag{34}$$
$$= f_k(\vec{x},t) + r_k(\vec{x},t),$$
$$(\vec{x},t) \in \Omega \times [t_{j-1}, t_j], \tag{35}$$
$$\mu_k(\vec{x},t)\frac{\partial \phi_k(\vec{x},t)}{\partial \vec{n}} + \beta(\vec{x},t)\phi_k(\vec{x},t) = g(\vec{x},t), \tag{36}$$
$$(\vec{x},t) \in (\delta\Omega \cap (\{x_k = 0\} \cup \{x_k = l_k\})) \times [t_{j-1}, t_j], \tag{37}$$
$$\phi_k(\vec{x},t_{j-1}) = \phi(\vec{x},t_{j-1}), \ j \geq 1, \ \gamma_k > 0, \ k = 1,2,3, \tag{38}$$
$$\sum_{k=1}^{3} f_k = f. \tag{39}$$

Reaction processes are simulated at the separate stage step.

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

Approximated solution of the direct problem in $t = t_j$ is evaluated as the average of the splitting stages

$$\phi(\vec{x},t_j) = \sum_{k=1}^{3} \gamma_k \phi_k(\vec{x},t_j) + \gamma_c \phi_c(\vec{x},t_j), \ \vec{x} \in \Omega. \tag{40}$$

As the temporal scales of convection-diffusion and chemical reaction processes are different a chemical step has its own temporal sub-grid on the interval $t_{j-1} < t < t_j$.

Illustration of the splitting solution is given on the Fig. 3. In the scenario a CO source has been placed in the middle of the spatial 2D domain, vertical wind velocities are 0.1, diffusion coefficient is 100, Square domain dimensions are $L_x = L_y = 50 * 350$. Number of spatial grid points $N_x = N_y = 50$. Time interval is $T = 24 * 60 * 60$.
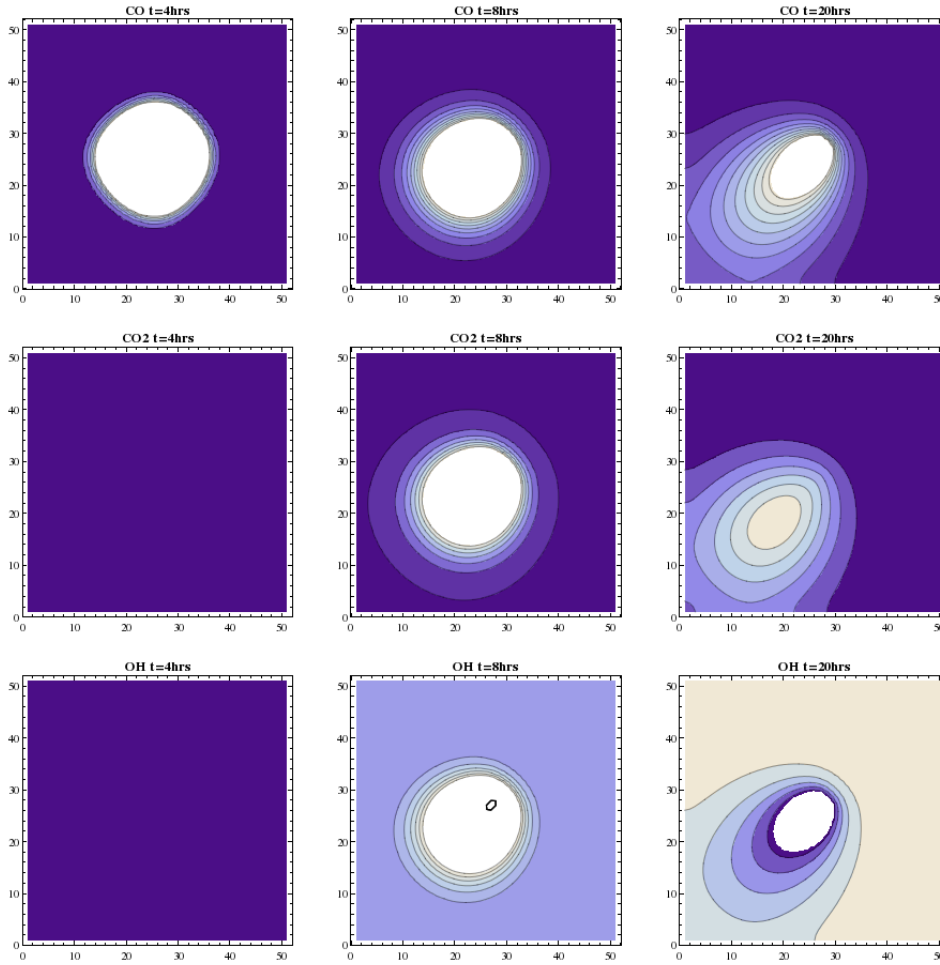
Figure 3. *2D Convection-diffusion-reaction simulation result for different time moments. In the scenario CO source has been placed in the middle of the domain, vertical wind velocities are **0.1**, diffusion coefficient is **100**, Square domain dimensions are $L_x = L_y = 50 * 350$, $T = 24 * 60 * 60$.*

3.5 Data assimilation for 1D convection-diffusion model

Nonstationary one-dimensional convection-diffusion model

$$\gamma_k \frac{\partial \vec{\phi}_k}{\partial t} + L_k \vec{\phi}_k = \vec{f_k} + \vec{r}_k, \quad \vec{\phi}_k(t^j) = \vec{\phi}(t^j), \quad k = 1,2,3.$$

is approximated on a spatial-temporal grid with upwind finite-difference scheme.

$$\gamma_k \frac{(\phi_k)_i^j - (\phi_k)_i^{j-1}}{\tau} + \left(L_k (\phi_k)^j \right)_i = (r_k)_i^j.$$

These equations can be presented in a tridiagonal form

$$-a_i\phi_{i+1}^j + b_i\phi_i^j = \phi_i^{j-1} + \tau r_i^j, i = 0, \tag{41}$$

$$-a_i\phi_{i+1}^j + b_i\phi_i^j - c_i\phi_{i-1}^j = \phi_i^{j-1} + \tau r_i^j, \ i = 1, \ldots, N-2, \tag{42}$$

$$b_i\phi_i^j - c_i\phi_{i-1}^j = \phi_i^{j-1} + \tau r_i^j, \ i = N-1, \tag{43}$$

where $\phi_i^j = (\phi_k)_i^j$ and N is the number of spatial grid points. For in situ measurement case we can introduce measurement system mask M_i^j that is equal to 1 in measurement point and 0 otherwise. In the points without measurements (where $M_i^j = 0$) we can set Ψ_i^j and σ_i to any value. In the case of stationary measurement posts we can modify our computational grid to combine grid points with measurement positions. Data assimilation problem solution is the minimum of the functional

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

wrt direct problem (41)-(43) where σ_i are standard deviation for the measurement device errors. Introducing Lagrange multipliers (adjoint functions) we obtain the augmented functional:

$$\bar{\Phi}(\phi^j, r^j, \phi^{*j}) = \left(\sum_{i=0}^{N-1}\left(\frac{\phi_i^j - \Psi_i^j}{\sigma_i}\right)^2 M_i^j + \alpha\sum_{i=0}^{N-1}(r_i^j)^2\right)\frac{\tau}{2}$$

$$+ \sum_{i=0}^{N-1}\left(-a_i\phi_{i+1}^j + b_i\phi_i^j - c_i\phi_{i-1}^j - \phi_i^{j-1} - \tau r_i^j\right)\phi_i^{*j}.$$

Equating first variations of the extended functional to 0, we obtain the following description of its stable point:

$$\partial_{\phi_i^*}\Phi(\phi^j, r^j, \phi^{*j}) = 0$$

is equivalent to the scheme of the direct problem (41)-(43).

$$\partial_{r_i^j}\Phi(\phi^j, r^j, \phi^{*j}) = 0$$

is equivalent to

$$\alpha r_i^j - \phi_i^{*j} = 0, \ i = 0, \ldots, N-1.$$

and

$$\partial_{\phi_i^j}\Phi(\phi^j, r^j, \phi^{*j}) = 0,$$

is equivalent to

$$-c_{i+1}\phi_{i+1}^{*j} + b_i\phi_i^{*j} = -\frac{M_i^j}{\alpha\sigma_i^2}(\phi_i^j - \Psi_i^j)\tau, \ i = 0,$$

$$-c_{i+1}\phi_{i+1}^{*j} + b_i\phi_i^{*j} - a_{i-1}\phi_{i-1}^{*j} = -\frac{M_i^j}{\alpha\sigma_i^2}(\phi_i^j - \Psi_i^j)\tau, i = 1, \ldots, N-2,$$

$$b_i\phi_i^{*j} - a_{i-1}\phi_{i-1}^{*j} = -\frac{M_i^j}{\alpha\sigma_i^2}(\phi_i^j - \Psi_i^j)\tau, \ i = N-1.$$

A system consisting of direct and adjoint problem can be assembled to the matrix equation [6,7],

$$-A_i\Phi_{i+1}^j + B_i\Phi_i^j = F_i^j,$$

$$-A_i\Phi_{i+1}^j + B_i\Phi_i^j - C_i\Phi_{i-1}^j = F_i^j,$$

$$B_i\Phi_i^j - C_i\Phi_{i-1}^j = F_i^j,$$

where

$$A_i = \begin{pmatrix} a_i & 0 \\ 0 & c_{i+1} \end{pmatrix}, \; B_i = \begin{pmatrix} b_i & -\tau \\ \dfrac{M_i^j \tau}{\alpha \sigma_i^2} & b_i \end{pmatrix}, \; C_i = \begin{pmatrix} c_i & 0 \\ 0 & a_{i-1} \end{pmatrix},$$

$$\Phi_i^j = \begin{pmatrix} \phi_i^{j+1} \\ \phi_i^{*j} \end{pmatrix}, \; F^j = \begin{pmatrix} \phi_i^{j-1} \\ \dfrac{M_i^j \tau}{\alpha \sigma_i^2} \Psi_i^j \end{pmatrix},$$

which is solved with the matrix sweep method. It can be easily seen that in case when all M_i are equal to zero data assimilation problem reduces to the direct problem.

3.6. Fine-grained approach to data assimilation

The main idea of the approach is to do data assimilation locally. Each split model step is substituted with the minimization problem [6,7]:

$$\bar{\Phi}_k(\phi_k^j, r_k^j, \phi_k^*) = \sum_{i=0}^{N_k-1} \left(\frac{(\phi_k^j)_i - \Psi_i^j}{\sigma_i} \right)^2 M_i^j \tau + \alpha_l \sum_{i=0}^{N_k-1} \left(r_{k_i}^j \right)^2 \tau$$
$$+ \left\langle \left(\left(I + \frac{\tau}{\gamma_k} L_k \right) \phi_k^j - \phi^{j-1} - \frac{\tau}{\gamma_k} r_k^j \right), \phi_k^* \right\rangle_l.$$
$$k = 1,2,3.$$

3.7. Data assimilation for CBM-IV system

Consider a CBM-IV system [2],[9],[5] of 20 reacting species over a square domain ($X = Y = 50 Intervals * 350$) for time period T=24*3600. Reaction rates depend on time (i.e. photochemical processes are considered). Three wind velocities are considered $(u,v) = (-0.6,0.6), (-6,-6), (0,0)$ and diffusion coefficient is $\mu = 100$. Background model is started with plausible constant concentrations. The perturbed model is used to simulate "measurement" data. In this model additional constant source emits in the middle of spatio-temporal domain. Emission rate has been chosen to provide 100 % relative difference between the perturbed and background model.

Species $NO2, SO2, CO, HCHO$ are assimilated to a background model in 9 points down the wind to the source on every time step.

In order to estimate the impact of Chemical model inclusion the following models has been run:
- **"Exact solution"** Convection-diffusion-reaction model with a constant source in the middle of the domain. Measurement data are taken from the model in a number of measurement points.
- **"Data assimilation"** These models has no source but can use measurement data to improve their solutions. These models start from the same conditions as the "Exact" one.
  - **"Passive tracers transport"** Convection-diffusion model without data assimilation.
  - **"Passive tracers transport + transport data assimilation"** Convection-diffusion model with data assimilation.
  - **"Passive tracers transport + reactions"** Convection-diffusion-reaction model without data assimilation.
  - **"Transport + transport data assimilation + reactions"** Convection-diffusion-reaction model without data assimilation of convection-diffusion part.
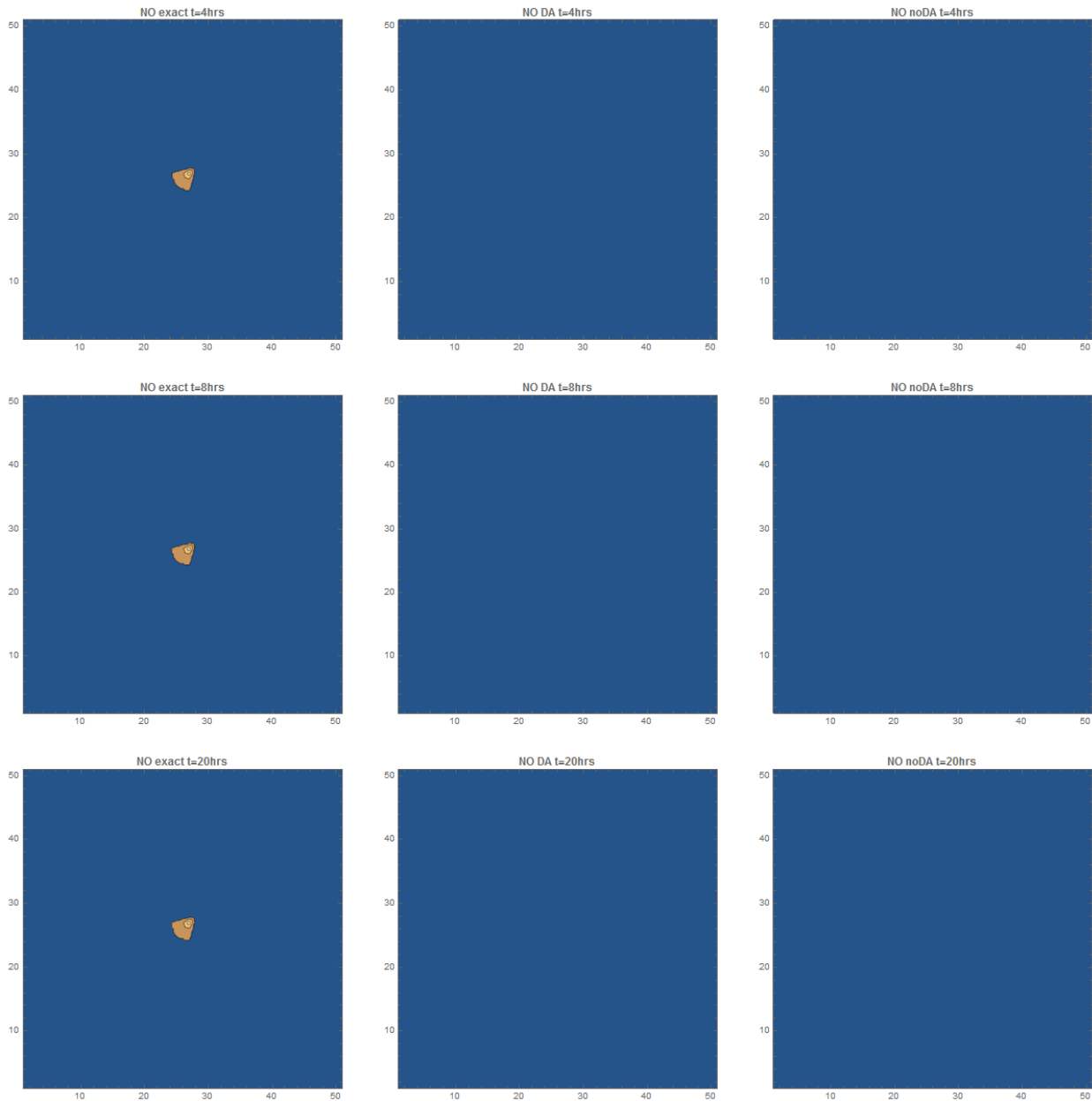
Figure 4. *Data assimilation result for a substance that is emitted in the "exact" problem for 3 time moments from top to bottom. "Exact" solution is on the left column, data assimilation result is on the center column and background model without DA in on the right column.*
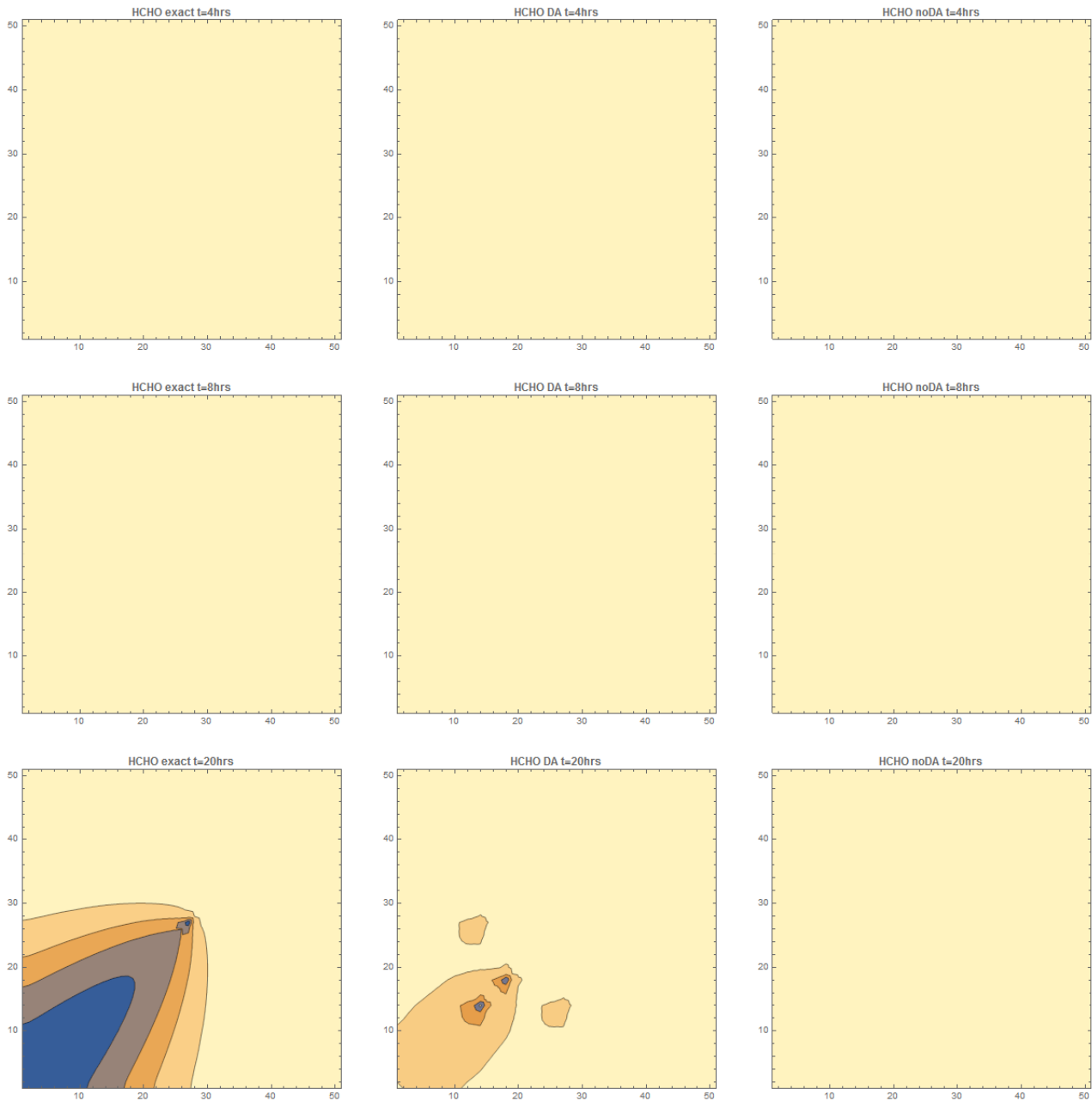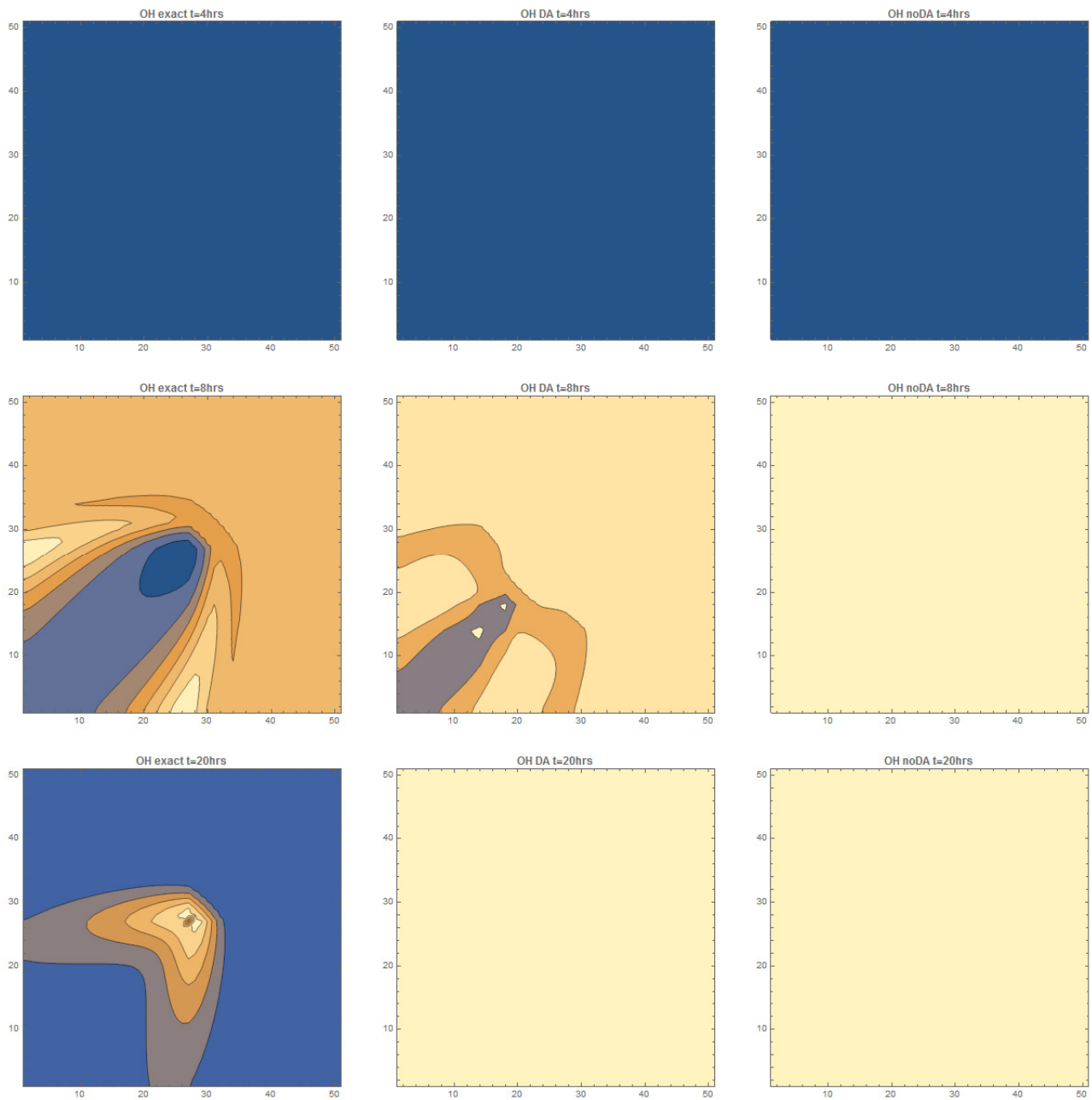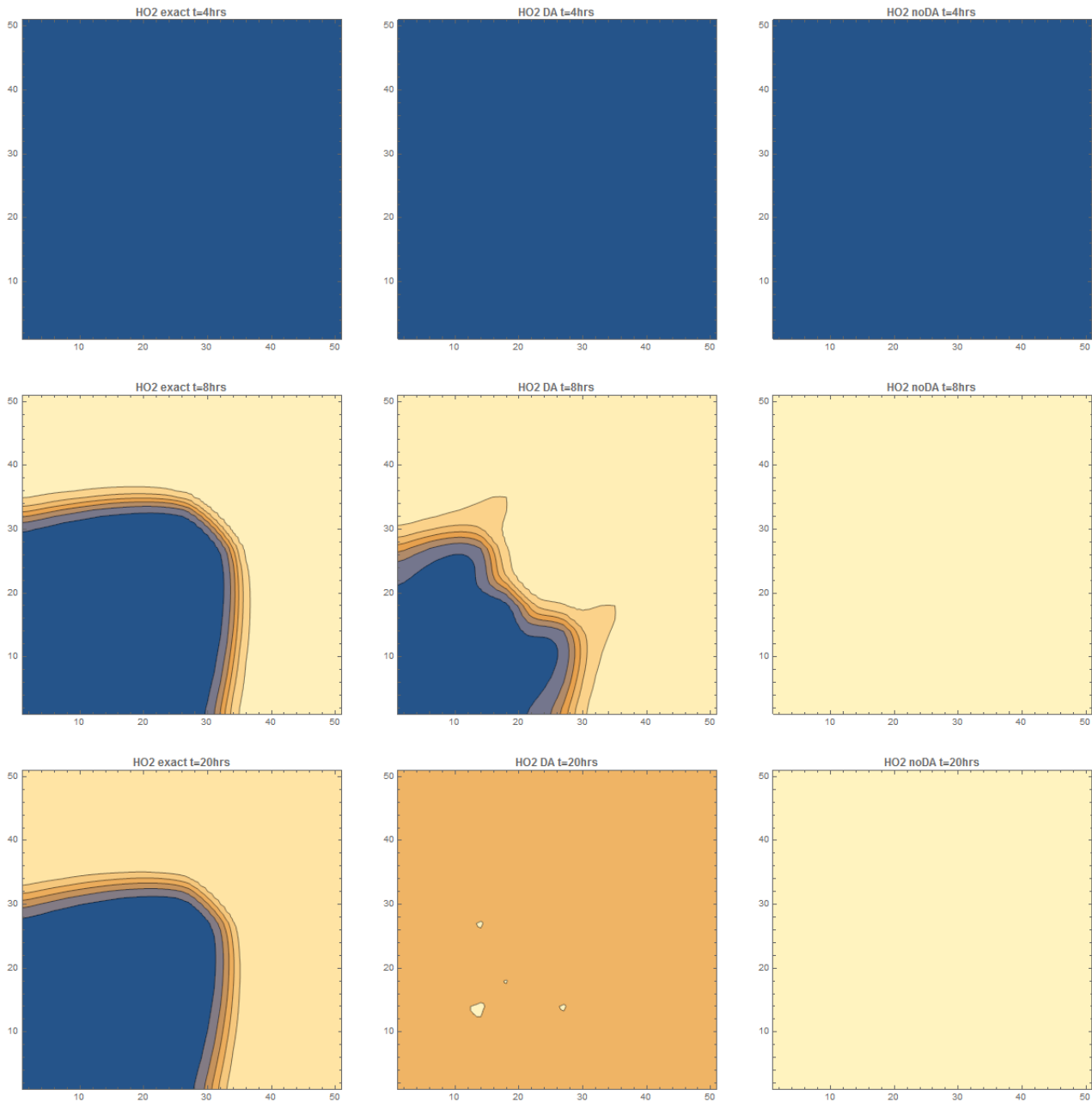
Figure 5. *Data assimilation result for a substance that is measured for 3 time moments from top to bottom. "Exact" solution is on the left column, data assimilation result is on the center column and background model without DA in on the right column.*
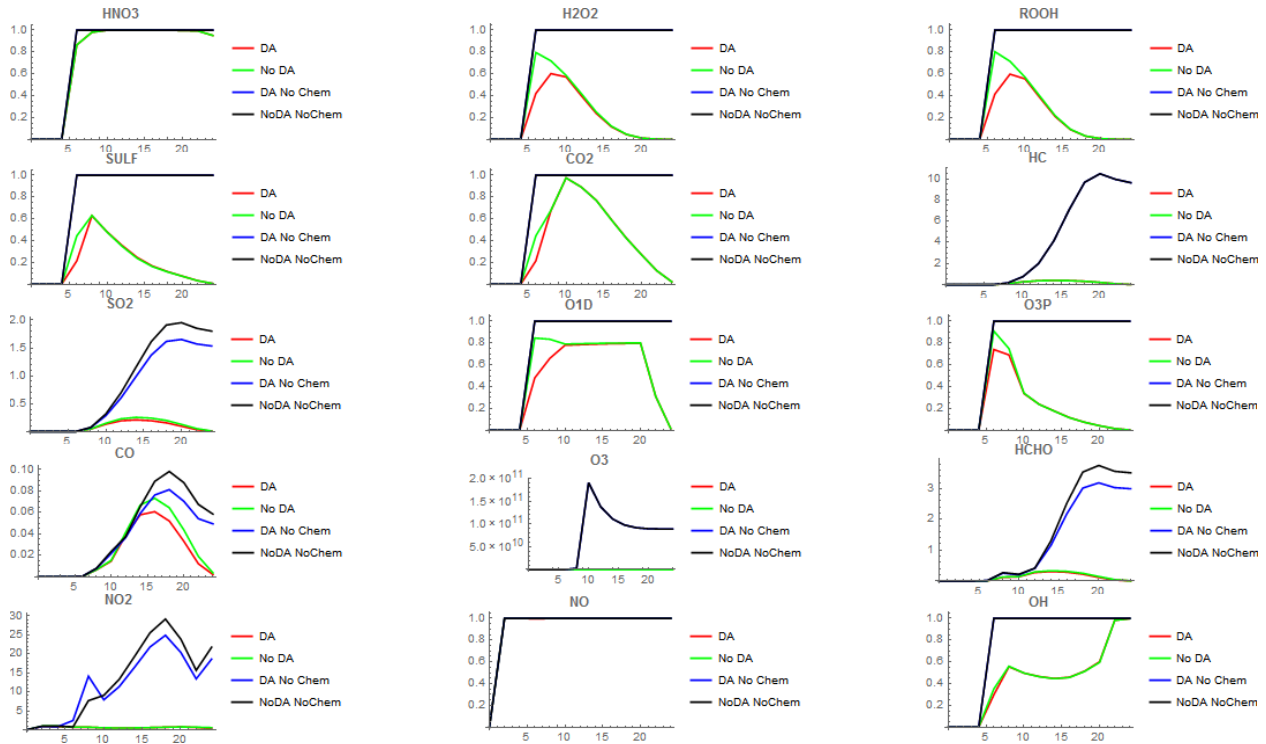
Figure 6. Data assimilation result for a substance that is neither measured nor emitted for 3 time moments from top to bottom. "Exact" solution is on the left column, data assimilation result is on the center column and background model without DA in on the right column.

Figure 7. *Data assimilation result for a substance that is neither measured nor emitted for 3 time moments from top to bottom. "Exact" solution is on the left column, data assimilation result is on the center column and background model without DA in on the right column.*

Figure 8. *Relative error with respect to time. Initial error growth is due to the beginning of photochemical processes (4-5'o clock am). In the legend: DA stands for Convection-diffusion-reaction model with Data assimilation, No DA stands for Convection-diffusion-reaction model without Data assimilation, DA No chem stands for Convection-diffusion model with Data assimilation, No DA No Chem stands for Convection-diffusion model without Data assimilation*
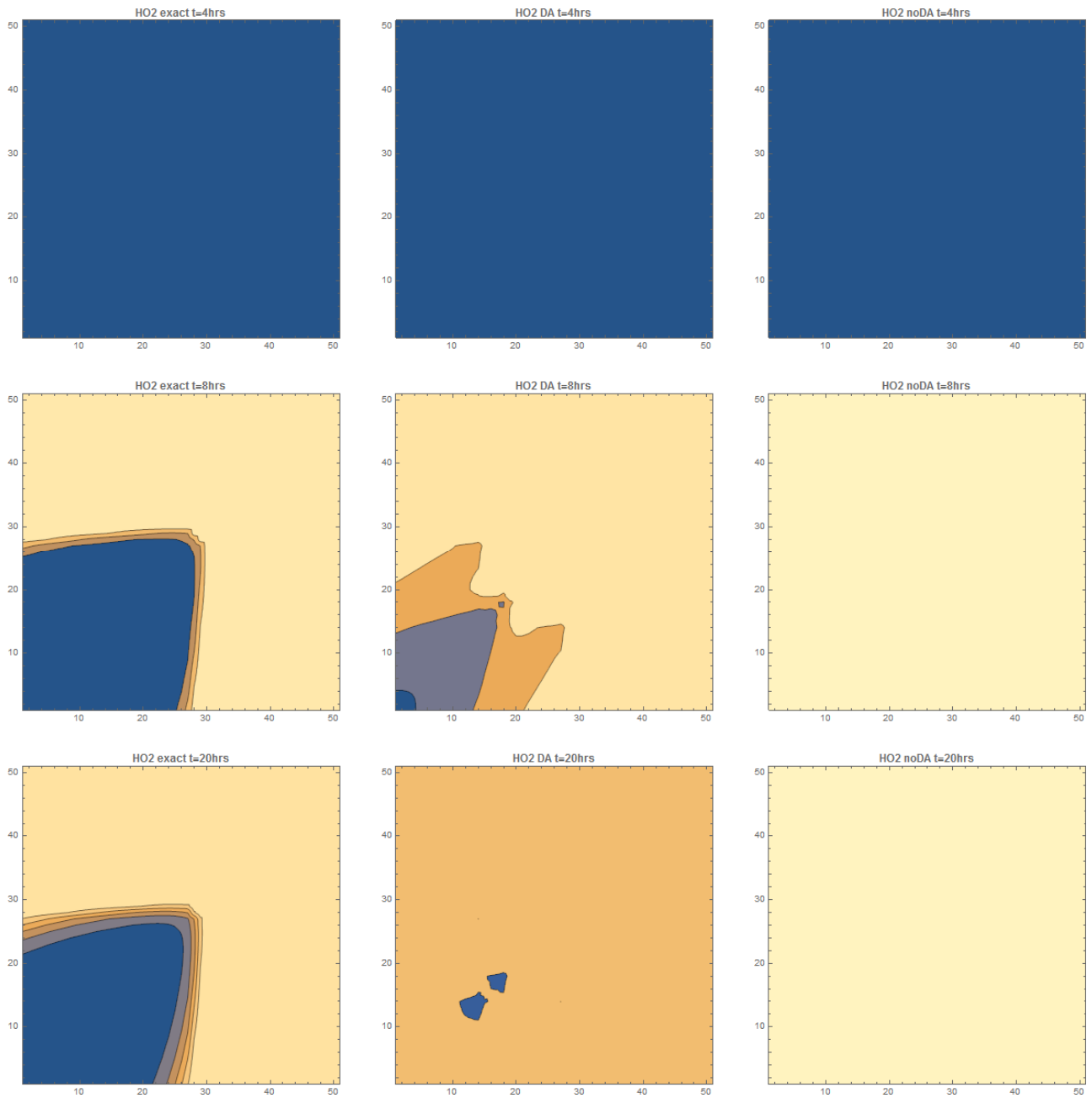
Figure 9. *"Strong wind" Data assimilation result for a substance that is neither measured nor emitted for 3 time moments from top to bottom. "Exact" solution is on the left column, data assimilation result is on the center column and background model without DA in on the right column.*
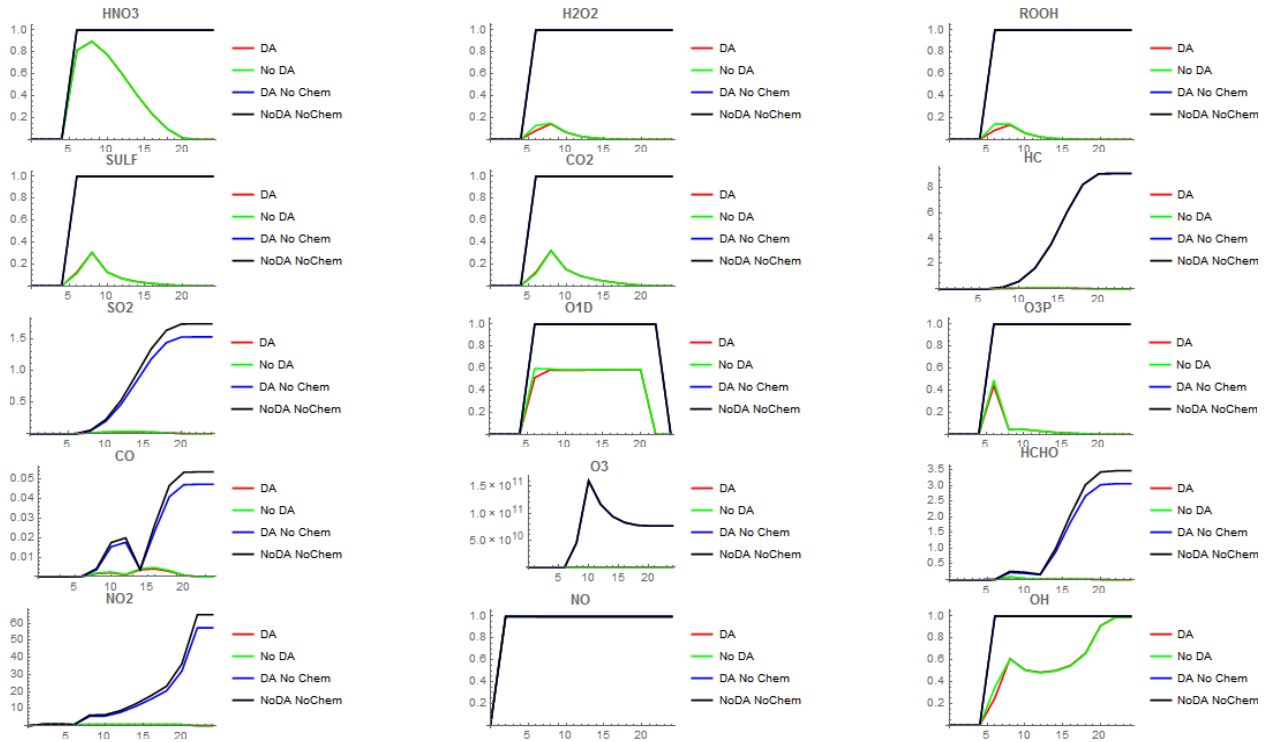
Figure 10. *"Strong wind" Relative error with respect to time. Initial error growth is due to the beginning of photochemical processes (4-5'o clock am). In the legend: DA stands for Convection-diffusion-reaction model with Data assimilation, No DA stands for Convection-diffusion-reaction model without Data assimilation, DA No chem stands for Convection-diffusion model with Data assimilation, No DA No Chem stands for Convection-diffusion model without Data assimilation.*
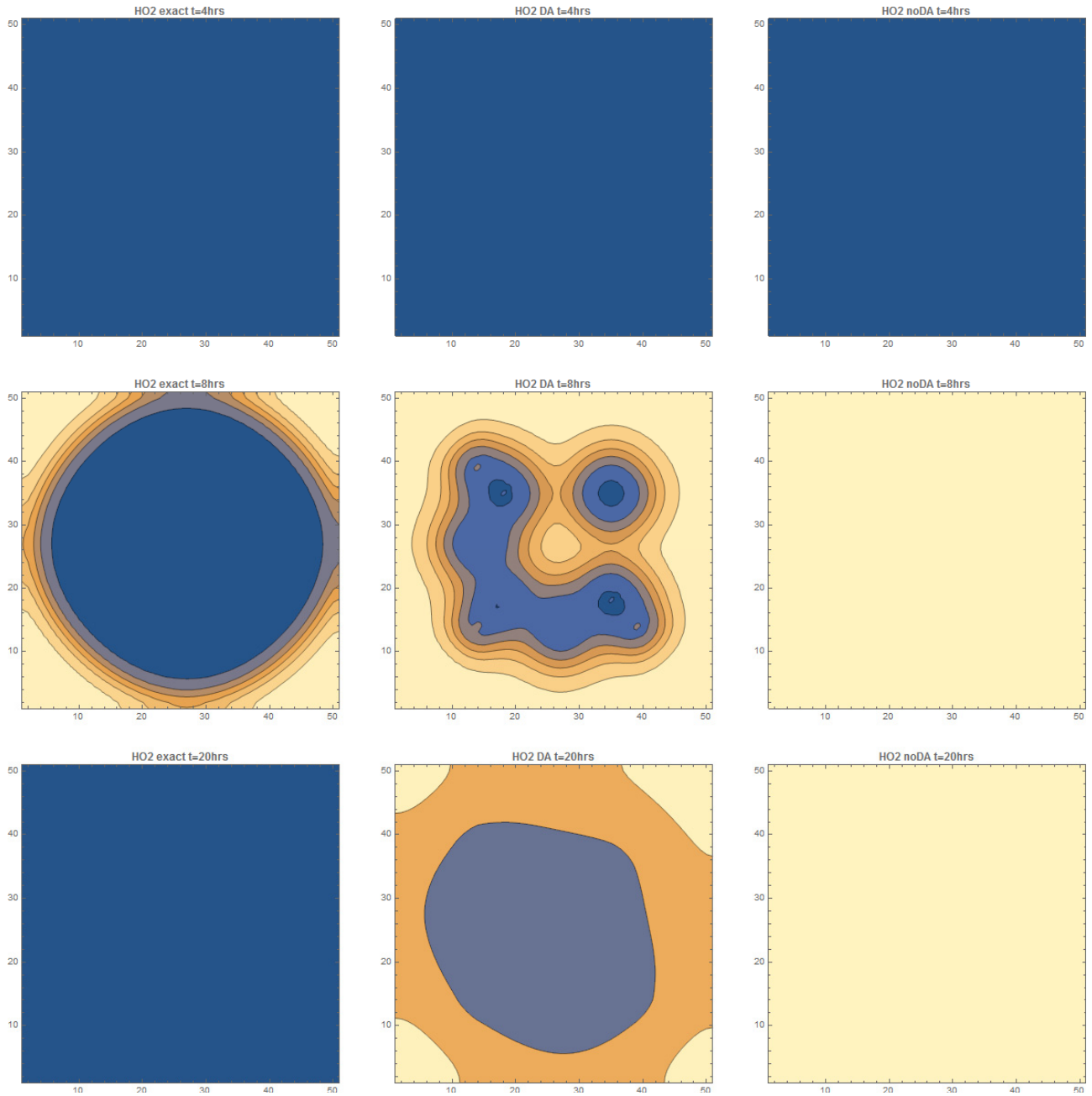
Figure 11. *"Strong wind" Data assimilation result for a substance that is neither measured nor emitted for 3 time moments from top to bottom. "Exact" solution is on the left column, data assimilation result is on the center column and background model without DA in on the right column.*
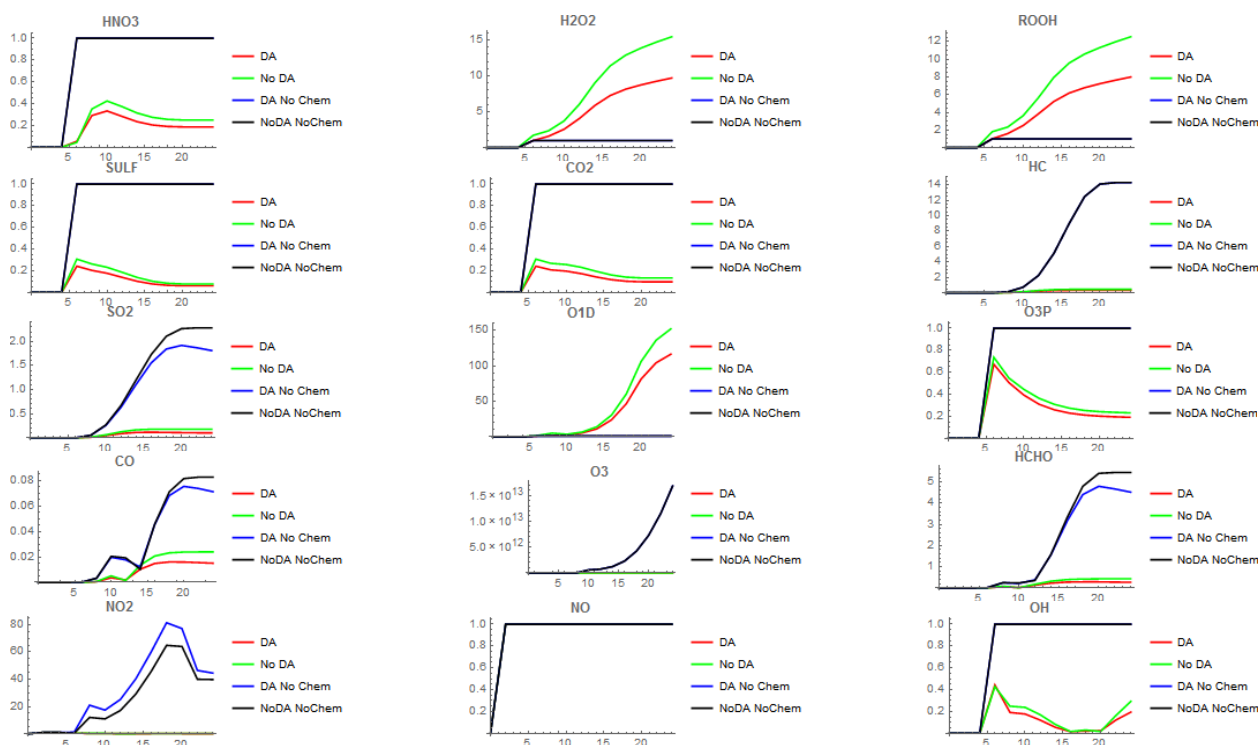
Figure 12. *"Calm". Relative error with respect to time. Initial error growth is due to the beginning of photochemical processes (4-5'o clock am). In the legend: DA stands for Convection-diffusion-reaction model with Data assimilation, No DA stands for Convection-diffusion-reaction model without Data assimilation, DA No chem stands for Convection-diffusion model with Data assimilation, No DA No Chem stands for Convection-diffusion model without Data assimilation.*

In the case of "Calm" wind (Fig. 12) chemistry model impact has been the most influential. In this case we can see the shift of solutions that can be seen with big relative errors ($H2O2, ROOH, O1D$) with respect to tracer-only models (black and blue lines). In these cases data assimilation were able to improve the situation (red and green lines). In the case of "Strong wind"(Fig. 10) the impact of the chemical model inclusion was also very important but data assimilation has very small influence (red and green lines almost coincide).

4. Conclusions

Data assimilation problems are solved with incomplete data. Lack of data is compensated by mathematical models that act as natural Tikhonov regularizers. Discrete-analytical schemes can be used to produce positive solutions for stiff systems of chemical kinetics equations. Combining splitting schemes and data assimilation schemes allows to construct computationally effective algorithms for data assimilation of contact measurements to convection-diffusion models. Operability of the developed algorithms has been evaluated by numerical experiments.

5. Acknowledgements

References

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

[2] M.W. Gery, G.Z. Whitten, J.P. Killus, and M.C. Dodge. A photochemical kinetics mechanism for urban and regional scale computer modeling. *Journal of Geophysical Research*, D 94:12925–12956, 1989.

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

[4] G. I. Marchuk. *Methods of Computational Mathematics*. Nauka, Moscow, 1980.

[5] R. et. al Nuterman. High resolution forecast meteorology and chemistry for a danish urban area. In *EMS*, 2011.

[6] V.V. Penenko. Variational data assimilation methods and inverse problems to study atmosphere, ocean and environment. *Sib. Zhur. Comp. Math.*, 12(4):341–351, 2009.

[7] A.V. Penenko and V.V. Penenko. Direct data assimilation method for convection-diffusion models based on splitting scheme. *Computational technologies*, 19(4):69–83, 2014.

[8] V. V. Penenko and E. A. Tsvetova. Variational methods for construction of monotone appoximations for atmospheric chemistry models. *Num. Anal. Appl.*, 6(3):210–220, 2013.

[9] W.R. Stockwell and W.G. Goliff. Comment on simulation of a reacting pollutant puff using an adaptive grid algorithm by r. k. srivastava et al. *JOURNAL OF GEOPHYSICAL RESEARCH*, 107(D22), 2002.

[10] B.E Shapiro, A. Levchenko, Meyerowitz E., B.J Wold, and E.D. Mjolsness. Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*, 19(5):677–678, 2003.

[11] A. Sandu and C. Tianfeng. Chemical data assimilation - an overview. *Atmosphere*, 2:426–463, 2011.

[12] A.A. Samarskii and P.N. Vabishchevich. *Computational Heat Transfer. Vol.1. Mathematical Modelling*. Wiley, Chichester, 1995.

[13] A.A. Samarskii and P.N. Vabishchevich. *Computational Heat Transfer. Vol.2. The Finite Difference Methodology*. Wiley, Chichester, 1995.

[14] A.N. Tikhonov, A.V. Goncharsky, V.V. Stepanov, and A.G. Yagola. *Numerical Methods for the Solution of Ill-Posed Problems*. Kluwer, Dordrecht, 1995.

[15] Hong Zhang, John C. Linford, Adrian Sandu, and Rolf Sander. Chemical mechanism solvers in air quality models. *Atmosphere*, 2(3):510–532, 2011.