
OPTICAL MODELS
AND DATABASES

Numerical Investigation of the Direct Variational Algorithm of Data Assimilation in the Urban Scenario

A. V. Penenko^{a,*}, Zh. S. Mukatova^{a,**}, V. V. Penenko^{a,***},
A. V. Gochakov^{b,****}, and P. N. Antokhin^{c,*****}

^a*Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

^b*Siberian Regional Hydrometeorological Research Institute, Novosibirsk, 630099 Russia*

^c*V.E. Zuev Institute of Atmospheric Optics, Siberian Branch, Russian Academy of Sciences, Tomsk, 634055 Russia*

**e-mail: a.penenko@yandex.ru*

***e-mail: zmukatova@yandex.ru*

****e-mail: penenko@sscc.ru*

*****e-mail: gochakov@sibnigmi.ru*

******e-mail: apn@iao.ru*

Received January 22, 2018

Abstract—The performance of a direct variational data assimilation algorithm with quasi-independent data assimilation at individual steps of the splitting scheme has been studied in a realistic scenario of air pollution assessment in the city of Novosibirsk by monitoring system data. For operation under conditions of a sparse monitoring network, an algorithm with minimization of the spatial derivative of the uncertainty (control) function adjusted to data assimilation is proposed. The use of the spatial derivative minimization increases the smoothness of the uncertainty (control functions) reconstructed, which has a positive effect on the reconstruction quality in the scenario considered.

Keywords: data assimilation, variational approach, splitting scheme, smart city

DOI: 10.1134/S102485601806012X

INTRODUCTION

It is important for the Smart City control system to include instruments that allow one to rapidly reconstruct and supplement information about the spatial and temporal distribution of contaminant fields in the city area and then forecast their changes. This is necessary due to the wide spectrum of gas and aerosol contaminants of anthropogenic origin that are injurious to human health (see, e.g., [1]). The main instrument allowing one to obtain information about the distribution and dynamics of contaminants in the city is mathematical simulation with rapid assimilation of contamination level measurement data obtained at stationary or mobile monitoring stations. The first results of simulating the processes of transport and transformation of atmospheric contaminants in urban conditions, with the example of Novosibirsk, were presented in [2].

Present-day models of transport and transformation of contaminants in the atmosphere take into account many various processes; for this reason, dimensions of state functions in those models can reach 10^{12} . Working with such dimensions creates spe-

cial requirements for data assimilation algorithms capable of operating in real time. In connection with this, direct (noniterative) algorithms of data assimilation are of special interest. A survey of such algorithms can be found in [3].

Let us consider a family of variational data assimilation algorithms in which the same data set is assimilated quasi-independently at individual stages of the splitting scheme. In this process, the constrained minimum of the objective functional is found at each stage on constraints of the mathematical model by the direct algorithm. The functional contains a discrepancy between the measured values and their modeled analogs, as well as a stabilization term including the norm of the uncertainty (control) function [4, 5]. An example of the operation of such an algorithm in a realistic scenario for Novosibirsk was presented in [6]. It was noted in the discussion that the exact and reconstructed solutions coincided qualitatively, but their absolute values were significantly different. We believe that this is caused by three factors. First, this is the character of the stabilization term in the objective functional of the variational data assimilation algorithm; it regulates the norm of the uncertainty func-

tion (additional source). Second, concentration fields of a widespread contaminant of the city atmosphere, e.g., SO₂ or CO, are not localized in small vicinities of monitoring stations (as it can occur when a cloud of a toxic agent moves over the city after a point-source release). Third, monitoring stations are situated at large distances from each other, which does not allow one to cover the area by measurement data for a good quantitative coincidence of the exact solution and the solution reconstructed by the abovementioned algorithm.

In [7], a direct algorithm of variational data assimilation with a stabilization term contained in the objective functional and regulating the norm of the derivative of the uncertainty function was presented. Solutions that can be obtained with the use of such a stabilization term are less localized but still agree with measurement data.

The aim of this work is to estimate the efficiency of using the direct algorithm of variational data assimilation with stabilization terms including both the norm of the uncertainty function itself and the norm of its spatial derivative in a realistic scenario for Novosibirsk.

1. FORMULATION OF THE PROBLEM

Let us consider the model of atmospheric transfer of contaminants [8–10]:

$$\frac{\partial \varphi(x, t)}{\partial t} + \operatorname{div}(\varphi(x, t)\mathbf{u}(x, t) - \mu(x, t)\operatorname{grad}\varphi(x, t)) = f(x, t) + r(x, t), \quad (x, t) \in D \times (0, T), \quad (1)$$

$$\mu(x, t) \frac{\partial \varphi(x, t)}{\partial n} = 0, \quad (x, t) \in \Gamma_{\text{out}}, \quad (2)$$

$$\varphi(x, t) = \varphi_b(x, t), \quad (x, t) \in \Gamma_{\text{in}},$$

$$\varphi(x, t) = \varphi_0(x), \quad x \in D, \quad t = 0. \quad (3)$$

Here, φ is the contaminant concentration, kg/m³; \mathbf{u} is the wind velocity vector, m/s; μ is the diffusion coefficient, m²/s; φ_b are background values of concentrations, kg/m³; φ_0 are the initial distributions of concentrations, kg/m³; f characterizes the power of the contaminant source, i.e., the mass of the emission per second for a cubic meter in the space; for this reason, its dimension in this formulation is kg/(m³ s); r corresponds to the uncertainty function added to the model for data assimilation; $D = (0, X) \times (0, Y)$, $X, Y > 0$ is a rectangular domain in the space; Γ_{out} is a portion of the boundary of the region $\partial D \times (0, T)$ on which the wind velocity is directed outwards from the region; Γ_{in} is a portion of the boundary on which the wind velocity is directed inwards to the region; and n is the external normal to the boundary of the region.

In the direct problem, μ , \mathbf{u} , φ_0 , φ_b , f , and r are known; it is required to find φ from (1)–(3). Assume that parameters of (1)–(3) are such that the direct

problem has a unique solution. Let the solution of the direct problem $\bar{\varphi}$ be obtained for a certain \bar{r} . Its noisy values are measured at M points of the space and time $\{(\chi_m, \theta_m)\}_{m=1}^M \subset D \times [0, T]$:

$$I_m = \bar{\varphi}(\chi_m, \theta_m) + \zeta_m, \quad m = 1, \dots, M, \quad (4)$$

where ζ_m is a perturbation of the m th measurement result.

In the data assimilation problem, μ , \mathbf{u} , φ_0 , φ_b , and f are given. For a certain time instant $t^* \in (0, T)$, let the measurement data I_m at time instants previous to t^* be available. According to this information, $\bar{\varphi}$ must be found from (1)–(4) for the whole interval $(0, T)$.

2. SOLUTION ALGORITHM

Let us define in the domain the temporal and spatial grids

$$\omega_T = \{t^k\}_{k=1}^{N_t} | 0 = t^1 < \dots < t^k < \dots < t^{N_t} = T\},$$

$$\omega_X = \{x_i\}_{i=1}^N | 0 = x_1 < \dots < x_i < \dots < x_N = X\},$$

$$\omega_Y = \{y_j\}_{j=1}^N | 0 = y_1 < \dots < y_j < \dots < y_N = Y\}.$$

For simplicity, let ω_T be a uniform grid with a step $\Delta \bar{t}$; ω_X and ω_Y be uniform grids with steps $\Delta x = \Delta y$, respectively. Consider an implicit scheme [11] approximating the boundary value problem (1)–(3) and additively averaged splitting scheme with respect to spatial variables [12]:

$$\phi^k = \{\varphi_0(x_i, y_j)\}_{i,j=1}^N, \quad k = 1, \quad (5)$$

$$\gamma_\xi \frac{\phi_\xi^k - \phi_\xi^{k-1}}{\Delta \bar{t}} = \Lambda_\xi^k \phi_\xi^k + F_\xi^k + r_\xi^k, \quad (6)$$

$$k = 2, \dots, N_t, \quad \xi \in \Xi;$$

$$\phi^k = \sum_{\xi \in \Xi} \gamma_\xi \phi_\xi^k, \quad \sum_{\xi \in \Xi} \gamma_\xi = 1, \quad \gamma_\xi \geq 0, \quad \Xi = \{x, y\}. \quad (7)$$

Here, ϕ , $\{\phi_\xi\}_{\xi \in \Xi}$, $\{r_\xi\}_{\xi \in \Xi} \in \mathbb{R}^{N \times N \times N_t}$ are vectors with elements $\{\phi_{ij}^k\}_{i,j,k=1}^{N,N,N_t}$, $\{(\phi_\xi^k)_{ij}\}_{i,j,k=1}^{N,N,N_t}$, and $\{(r_\xi^k)_{ij}\}_{i,j,k=1}^{N,N,N_t}$; $\mathbf{F}_\xi \in \mathbb{R}^{N \times N \times N_t}$ are vectors containing elements $\{f_{ij}^k\}_{i,j,k=1}^{N,N,N_t}$, $f_{ij}^k = f(x_i, y_j, t^k)$, and corresponding aggregates from boundary conditions for this splitting scheme. Suppose that measurement points coincide with grid point $\{(\chi_m, \theta_m)\}_{m=1}^M \subset \omega_X \times \omega_Y \times \omega_T$.

For the solution of the problem of variational data assimilation for a temporal step k and stage of splitting along the x axis, we consider points of constrained minima $(\check{\phi}_x)_j^k$ of the objective functionals J_{xj}^k at $j = 1, \dots, N$ (for brevity, we denote $\phi_i^k = (\phi_x)_i^k$,

$r_i^k = (r_x)_{ij}^k$, $F_i^k = (F_x)_{ij}^k$, $H_i^k = H_{ij}^k$, $I_i^k = I_{ij}^k$, and $\Delta t = \Delta \bar{t} / \gamma_\xi$:

$$J_{xj}^k(\phi^k, r^k) = \left(\sigma \sum_{i=1}^N (\phi_i^k - I_i^k)^2 H_i^k + \rho \sum_{i=1}^N (r_i^k)^2 + \eta \sum_{i=1}^{N-1} \left(\frac{r_i^k - r_{i+1}^k}{\Delta x} \right)^2 \right) \Delta t, \quad (8)$$

for which ϕ^k and r^k are related by Eq. (6) (at $\xi = x$); H_i^k is the mask of the observational system: $H_i^k = 1$ at the observation point $\{(\chi_m, \theta_m)\}_{m=1}^M \subset \omega_X \times \omega_Y \times \omega_T$ and 0 in the remaining part of the region $\omega_X \times \omega_Y \times \omega_T$. Expression (6) is reduced to the tridiagonal numerical scheme

$$-L_i \phi_{i-k}^k + C_i \phi_i^k - R_i \phi_{i+1}^k = \phi_i^{k-1} + \Delta t F_i^k + \Delta t r_i^k, \quad i = 1, \dots, N.$$

Then, the proposition from [7] is valid: the constrained minimum point of functional (8) on constraints (6) is calculated in the process of solving the matrix system

$$-\bar{L}_i \Phi_{i-1} + \bar{C}_i \Phi_i - \bar{R}_i \Phi_{i+1} = \bar{F}_i, \quad i = 1, \dots, N, \quad (9)$$

where

$$\begin{aligned} \bar{R}_i &= \begin{cases} \begin{pmatrix} R_i & 0 & 0 \\ 0 & L_{i+1} & 0 \\ 0 & 0 & \frac{2\eta\Delta t}{\Delta x^2} \end{pmatrix}, & i = 1, \dots, N-1, \\ 0, & i = N; \end{cases} \\ \bar{L}_i &= \begin{cases} 0, & i = 1, \\ \begin{pmatrix} L_i & 0 & 0 \\ 0 & R_{i-1} & 0 \\ 0 & 0 & \frac{2\eta\Delta t}{\Delta x^2} \end{pmatrix}, & i = 2, \dots, N; \end{cases} \\ \bar{C}_i &= \begin{pmatrix} C_i & 0 & -\Delta t \\ 2\sigma H_i^k \Delta t & C_i & 0 \\ 0 & -\Delta t & 2\rho\Delta t + \frac{p_i^C \eta \Delta t}{\Delta x^2} \end{pmatrix}, \quad i = 1, \dots, N, \\ p_i^C &= \begin{cases} 2, & i \in \{1, N\}, \\ 4, & 2 \leq i \leq N-1; \end{cases} \\ \bar{F}_i &= \begin{pmatrix} \phi_i^{k-1} + \Delta t F_i^k \\ 2\sigma H_i^k I_i^k \Delta t \\ 0 \end{pmatrix}, \quad \Phi_i = \begin{pmatrix} \phi_i^k \\ \psi_i \\ r_i^k \end{pmatrix}, \quad i = 1, \dots, N. \end{aligned}$$

To solve (9) without iterations, one can use a matrix analog of the sweep method. Similar operations are carried out for other splitting stages, which results in the estimates $\{\check{\phi}_\xi\}_{\xi \in \Xi}$. For the solution of the problem of data assimilation at a step k , we consider

$$\check{\phi}^k = \sum_{\xi \in \Xi} \gamma_\xi \check{\phi}_\xi^k. \quad (10)$$

3. PREPARATION OF THE SCENARIO FOR MODELING

Preparation of coefficients for the contaminant transfer model (5)–(7) as a component of the algorithm for solving the inverse problem involved meteorological quantities calculated using the WRF-Chem model [13] in an area bounded by coordinates 54.75°–55.16° N, 82.66°–83.37° E (99 × 99 points) with a horizontal grid step of 460 m containing 30 vertical levels up to the level of 50 hPa. The first level of the model corresponds on average to 28.7 m from the Earth's surface. An adaptive temporal step from 1 to 20 s was specified.

From the horizontal wind velocity components (U and V), the wind velocity vector \mathbf{u} was formed by averaging over first eight vertical WRF-Chem levels. Similarly, from the vertical diffusion coefficient (EXCH_H), the coefficient μ was formed by averaging over the first eight WRF-Chem levels. To prepare files of coefficients, calculations for the model period from 0000 UT of July 13 to 0000 UT of July 17, 2008, were carried out. The initial and background values ϕ_0 and ϕ_b were chosen to be zero. Parameters of grids in the experiments with data assimilation are as follows: $N = 99$, $N_t = 57600$, and $T = 96 \times 3600$.

Suppose that the contamination sources are represented by objects of two types: a point source with time-constant power and a point source with a given time variation of the power. The former model emissions from combined heat and power plants and other contaminants; for brevity we below call them CHPP. The latter model emissions from transport, and their power depends on the traffic density which, in turn, depends on time. For brevity, we call the second sources roads. In the discrete system, point sources are modeled using Kronecker delta functions δ_{ij} :

$$\begin{aligned} \delta(i - \bar{i}, j - \bar{j}) &= \frac{\delta_{i\bar{i}} \delta_{j\bar{j}}}{\delta x_i \delta y_j}, \quad \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \\ \delta x_i &= \begin{cases} \Delta x / 2, & i = 1, \\ \Delta x, & 1 < i < N, \\ \Delta x / 2, & i = N, \end{cases} \end{aligned}$$

where $\{\bar{i}, \bar{j}\}$ is the index of location of the given source and δy_j is defined by analogy with δx_i . Let $\{\bar{i}_s, \bar{j}_s\}_{s=1, \dots, \bar{N}}$ denote indices of nodes for constant sources and



Fig. 1. Emission sources (triangles) and monitoring stations (circles) (large triangles with vertices upwards mark CHPPs; small triangles with vertices downwards mark roads).

$\{\tilde{i}_s, \tilde{j}_s\}_{s=1, \dots, \tilde{N}}$ denote indices of nodes for sources variable in time. In this case, the source function

$$\begin{aligned} \bar{r}_{ij}^k = & \sum_{s=1}^{\tilde{N}} \bar{q}_s \delta(i - \bar{i}_s, j - \bar{j}_s) \\ & + \sum_{s=1}^{\tilde{N}} \tilde{q}_s Q(t^k) \delta(i - \tilde{i}_s, j - \tilde{j}_s) \end{aligned} \quad (11)$$

(\bar{q}_s and \tilde{q}_s are powers of sources of the constant and variable types, respectively, kg/(m s)). The dimensionless parameter $Q(t)$ describes the a priori given time variation of the source intensity. We suppose that the a priori information about the sources when solving the problem of data assimilation is absent, i.e., $\mathbf{F}_\xi^k = 0$.

The information about the arrangement of sources was taken from data about city CHPPs and boiler houses, as well as the road map with statistically processed data about total traffic congestion during the period 2015–2017. Since the processed data are a highway congestion function of time, they slightly depend on the total quantity of transport facilities and can be interpreted as averaged information about the rhythm of motion in the agglomeration.

The road map was taken from Static API [14] of the Yandex service (road traffic) [15] and georeferenced. Road sections referenced to the emission grid of the calculation domain are marked in Fig. 1 by black triangles with vertices downwards. The parameter $Q(t)$

was specified using half-hour estimates of the relative quantity of red and yellow pixels on the map of the service; the pixels denoted busy sections of roads. The period of $Q(t)$ corresponds to the following days of the week: Sunday, Monday, Tuesday, and Wednesday of every week of July, and reflects the traffic dynamics in days of model calculations. The calculation result for the traffic intensity $Q(t)$ is presented in Fig. 2.

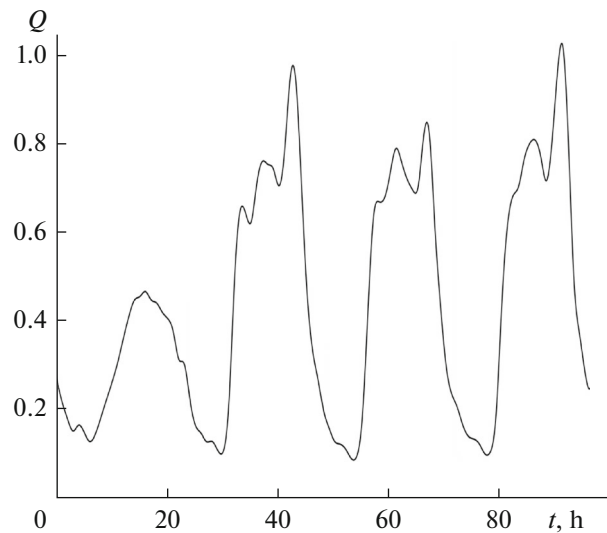


Fig. 2. Temporal variation of the power Q of road sources.

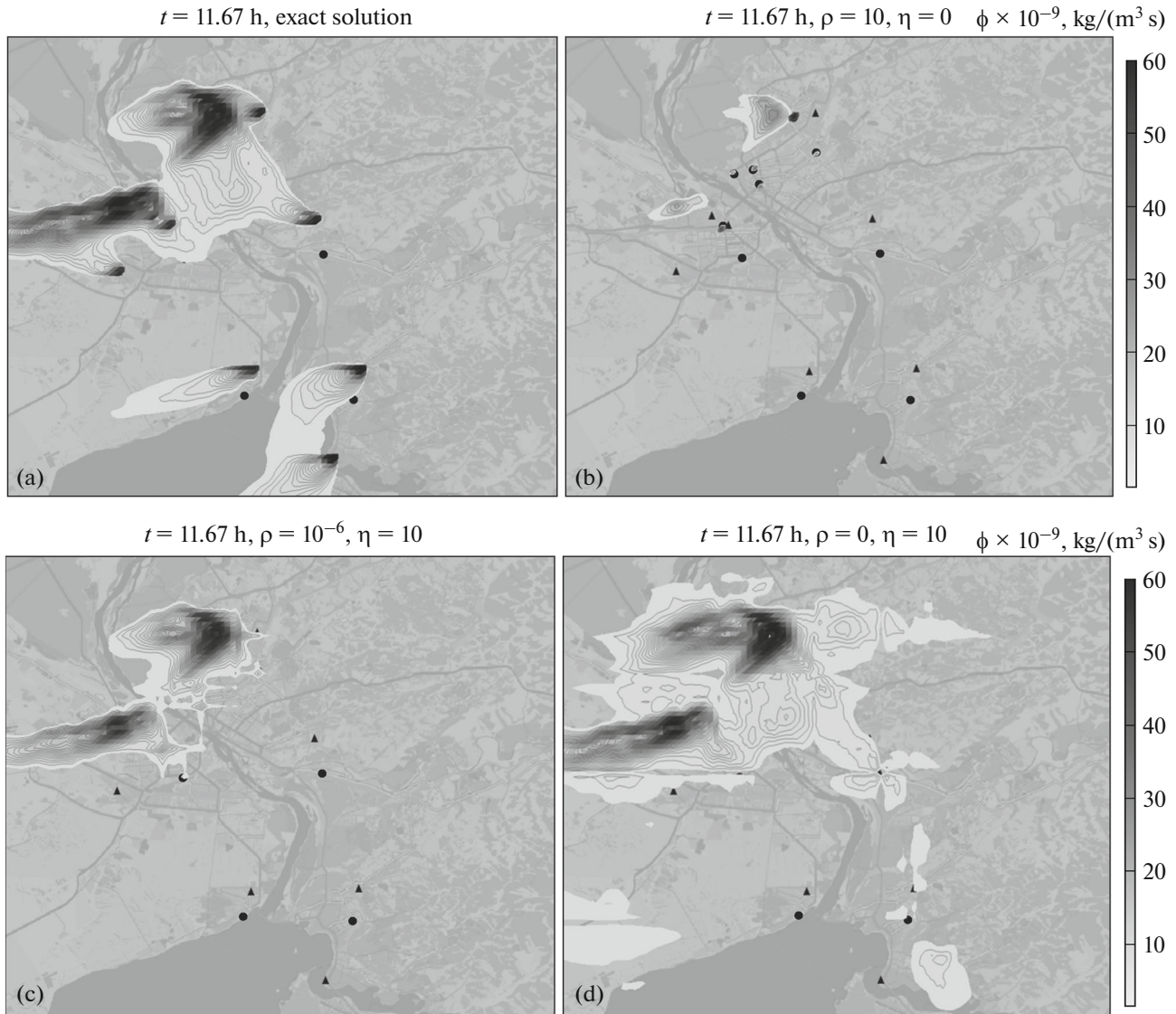


Fig. 3. Results of reconstructing the concentration field: (a) exact solution at different values of regularization parameters (b–d) at a given time instant of measurement data assimilation; concentrations above $1.5 \times 10^{-9} \text{ kg}/(\text{m}^3 \text{s})$ are shown.

Let us construct a realistic scenario according to Fig. 1. Let the total power of all sources in parameterization (11) be equal to $10^{-9} \text{ kg}/(\text{m s})$ and the total power of roads be equal to that of CHPPs:

$$\sum_{s=1}^{\bar{N}} \bar{q}_s + \sum_{s=1}^{\tilde{N}} \tilde{q}_s = 10^{-9}, \quad \sum_{s=1}^{\bar{N}} \bar{q}_s = \sum_{s=1}^{\tilde{N}} \tilde{q}_s.$$

In this case, the power of each of eight CHPPs is equal to $0.0625 \times 10^{-9} \text{ kg}/(\text{m s})$, and the power of each of 337 roads is $0.0015 \times 10^{-9} \text{ kg}/(\text{m s})$. The direct problem is solved with these exact values of the source $\bar{\mathbf{r}}$. From the obtained exact solution $\bar{\phi}$ at given points of measurement sites according to [16] (circles in Fig. 1),

values for using as $\bar{\mathbf{I}}$ are chosen every 20 min of the model time. The problem of data assimilation is solved with these measurement data at an unknown $\bar{\mathbf{r}}$; as a result, an estimate of concentration fields ϕ is obtained.

4. RESULTS OF DATA ASSIMILATION

The measurement weight is $\sigma = 10^{10}$ in all presented calculations. Let us study the dependence of the result on the relation between the regularization parameters ρ and η . Figure 3 shows results of data assimilation at the instant of the arrival of measurement results. One can note that the solution in Fig. 3b without minimization of the derivative is more local-

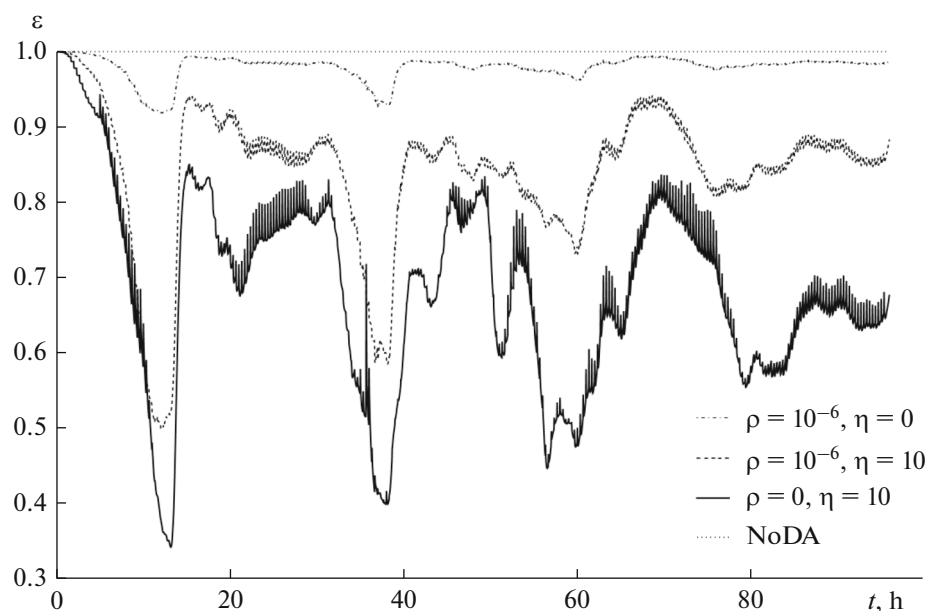


Fig. 4. Relative error ε of the concentration field reconstruction as a function of time for different values of regularization parameters.

ized in the vicinity of measurement points as compared to solutions in Figs. 3c and 3d. In the last figures, cross-shaped features of the solution with centers at the measurement points are visible. This is a consequence of quasi-independent data assimilation at individual stages of the splitting scheme. In future work, it is planned to reduce the effect of this phenomenon on the solution.

Let us estimate the solution accuracy quantitatively. Figure 4 presents the relative root-mean-square error of the solution of the data assimilation problem as a function of time for different regularization parameters.

In the case without data assimilation (NoDA), the relative error is equal to 1. The solution for assimilation (regularization) parameters $\rho = 0$ and $\eta = 10$ (corresponding to Fig. 3d) has the minimum error among those presented.

CONCLUSIONS

The efficiency of the direct algorithm of variational data assimilation with quasi-independent data assimilation at individual steps of the splitting scheme has been studied in a realistic scenario of estimating the level of atmospheric contamination in Novosibirsk from measurements of a monitoring system. Operation of the algorithm with minimization of the spatial derivative of the uncertainty (control) function in the objective functional, due to which the data are assimilated, has been considered under conditions of a sparse monitoring network. This increased the smoothness of the reconstructed uncertainty func-

tions, which had a positive effect on the reconstruction quality in the scenario under consideration.

ACKNOWLEDGMENTS

The new modification of the direct data assimilation algorithm was carried out under state contract no. 0315-2016-0004 for the Institute of Computational Mathematics and Mathematical Geophysics. Construction of the scenario for the simulation, adaptation of the algorithm to the conditions of Novosibirsk, and numerical experiments were supported by the Russian Foundation for Basic Research and Government of Novosibirsk oblast (project no. 17-41-543309).

REFERENCES

1. *Air Quality Guidelines Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide (A EURO Publication)* (WHO, Regional Office for Europe, Denmark, Copenhagen, 2006).
2. V. V. Penenko, A. E. Aloyan, N. M. Bazhin, and G. I. Skubnevskaya, "Numerical model for hydrometeorological conditions and pollution of cities and industrial regions," *Meteorol. Gidrol.*, No. 4, 5–15 (1984).
3. M. Bocquet, H. Elbern, H. Eskes, M. Hirtl, R. Zabkar, G. R. Carmichael, J. Flemming, A. Inness, M. Pagowski, J. L. Perez Camano, P. E. Saide, Jose R. San, M. Sofiev, J. Vira, A. Baklanov, C. Carnevale, G. Grell, and C. Seigneur, "Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models," *Atmos. Chem. Phys.* **14**, 32233–32323 (2014).
4. A. V. Penenko and V. V. Penenko, "Direct data assimilation method for convection-diffusion models based on

- splitting scheme,” *Vychisl. Tekhnol.* **19** (4), 69–83 (2014).
5. A. V. Penenko, V. V. Penenko, and E. A. Tsvetova, “Sequential data assimilation algorithms for air quality monitoring models based on a weak-constraint variational principle,” *Numer. Anal. Appl.* **9**, 312–325 (2016).
 6. V. V. Penenko, A. V. Penenko, and E. A. Tsvetova, “Variational approach to the study of processes of geophysical hydro-thermodynamics with assimilation of observation data,” *J. Appl. Mech. Tech. Phys.* **58**, 771–778 (2017).
 7. A. Penenko, V. Penenko, and Z. Mukatova, “Direct data assimilation algorithms for advection-diffusion models with the increased smoothness of the uncertainty functions,” in *2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), September 18–22, 2017, Novosibirsk*, p. 126–130.
 8. V. V. Penenko, *Numerical Simulation Techniques for Atmospheric Processes* (Gidrometeoizdat, Leningrad, 1981) [in Russian].
 9. G. I. Marchuk, *Simulation in Environmental Problems* (Nauka, Moscow, 1982) [in Russian].
 10. V. V. Penenko and A. E. Aloyan, *Models and Methods for Environmental Safety Problems* (Nauka, Novosibirsk, 1985) [in Russian].
 11. A. A. Samarskii, *Introduction in the Theory of Difference Schemes* (Nauka, Moscow, 1971) [in Russian].
 12. D. G. Gordeziani and G. V. Meladze, “Somulation of the third boudary problems for multi-dimensional parabolic equations in an arbitrary region by one-dimensional equations,” *Zhurn. Vychisl. Matem. Matem. Fiz.* **14**, 246–250 (1974).
 13. W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers, A Description of the Advanced Research WRF Version 3. NCAR/TN 475 + STR Technical Note, UCAR. <http://dx.doi.org/> (Cited January 19, 2018). doi 10.5065/D68S4MVH
 14. Yandex Static API. https://tech.yandex.ru/maps/doc/staticapi/1.x/dg/concepts/input_params-docpage/ (Cited January 19, 2018).
 15. <http://static-maps.yandex.ru/1.x/?ll=82.920430,55.030199&spn=0.31457,0.15&l=trf> (January 19, 2018).
 16. http://www.nso.ru/sites/test.new.nso.ru/wodby_files/files/wiki/2014/01/korrektura_gosdoklad-2015.compressed.pdf (January 19, 2018).

Translated by A. Nikol'skii