Air pollution modelling in urban environment based on a priori and reconstructed data

A V Gochakov¹, A V Penenko², P N Antokhin⁴, and A B Kolker¹,⁵

¹Siberian Regional Research Hydrometeorological Institute, 30, Sovetskaya, Novosibirsk 630099, Russia
²Institute of Computational Mathematics and Mathematical Geophysics SB RAS, 6, Akademika Lavrentieva, Novosibirsk, 630090, Russia
³Novosibirsk National Research State University, 2, Pirogova, Novosibirsk, 630090, Russia,
⁴Institute of Atmospheric Optics SB RAS, 1Akademika Zueva, Tomsk, 634055, Russia
⁵Novosibirsk State Technical University, 20, Prospekt K. Marksa, Novosibirsk, 630073, Russia

¹ gochakov@sibnigmi.ru

Abstract. This paper presents preliminary results of the effectiveness analysis of an air quality forecasting system for the city of Novosibirsk with replenishment of the missing information on emission sources by solving an inverse problem with urban monitoring network data. In solving the inverse problem, a priori information about the location and mode of the sources is used. To simulate concentration distributions, the WRF-Chem model is used, and a simplified model of chemical transport is applied to solving the inverse problem. These models are offline coupled in a hybrid forecast system in order to improve the initial information about the spatial distribution of emission intensity and air quality forecast, respectively. The results of numerical experiments and their analysis are presented. The influence of an urban parameterization on the results of the forecast is shown.

1. Introduction

Numerical methods of forecasting air pollution in cities are promising, intensively developing, and a relevant direction in meteorology and environmental protection. The range of negative influence of trace gases of anthropogenic origin on human health is rather wide (see, for example, [1]). Modern numerical weather prediction models which include the component of gases and aerosols distribution and transformation take into account a variety of a priori information that affects the prediction quality.

The meteorological regime of an urban agglomeration has differences from that of the natural environment. The urban environment can contribute to the accumulation of harmful impurities [2]. Usually, mesoscale weather prediction models do not include a detailed description of the processes of urban canyon circulation. Therefore, a class of urban canopy models was created to describe the urban processes and the influence of human activity on the city climatic conditions [3]. These parameterizations require specification of the parameters related to the classification of urban areas;
building size and location; surface characteristics; the percentage fraction of artificial constructions, natural landscape and vegetation.

When simulating the pollutant distribution in an urban agglomeration, clarification of the emissions (location of sources and intensity of their emissions) is of paramount importance [4][5][6]. Global and regional emission bases contain averaged data with a resolution of up to \(0.1^\circ \times 0.1^\circ\). The information included in these bases does not always correspond to the fact, which leads to significant discrepancies when comparing the modeling results with measurement data. Another source for emission correction is annual reviews of the state and pollution of the environment which often contain outdated, averaged, or approximate information.

Real-time pollution information is obtained from stations of urban air quality monitoring. These data can be used to determine the location and intensity of actual emission sources and restore the missing information by solving an inverse problem based on an appropriate mathematical model of chemical transport which binds the emission source functions with the pollutant concentration fields [7]. The accuracy of the solution also depends on \textit{a priori} data on the location of emission sources and their temporal characteristics.

This study deals with the issue of preparing \textit{a priori} and forecast information to solve the inverse problem of recovering the intensity of SO\textsubscript{2} emission sources for the city of Novosibirsk. The impact of urban parameterization on meteorological characteristics is studied. The results of the SO\textsubscript{2} distribution obtained by simulation with recovered data of emissions in the passive tracer mode are compared with those obtained from the EDGAR database. The results of simulation for urban and non-urban modeling scenarios are compared.

2. Description and configuration of the models

Offline coupling of the models was used in this study, including a chemical transport model developed at ICM&MG SB RAS within an Inverse Modeling and Data Assimilation Framework (IMDAF) and the open source WRF model with the Chem component [8]. The WRF model was used to calculate the meteorological elements (wind speed components, diffusion coefficients) required to assign the coefficients for the chemical transport model and also to build a retrospective forecast of the pollutant concentration field distribution based on the data on emission sources recovered by solving the inverse problem using the IMDAF model.

The calculations were carried out for two domains. The first domain (D1) had a horizontal grid step of 1380 m and included 30 vertical levels up to a level of 50 hPa. The second nested domain (D2) represented the Novosibirsk agglomeration and was limited to coordinates 54.75 - 55.16 °N. 82.66 - 83.37 °E. (99 x 99 points) with a horizontal grid step of 460 m and contained 30 vertical levels to a level of 50 hPa. To calculate the nested domain D2, we used the downscaling scheme "1-way nesting". The average height of the first model level corresponds to a height of 28.7 m from the Earth’s surface. An adaptive time step was used for the computational experiments. The minimum step was chosen equal to 1 second, and the maximum step was up to 20 seconds.

The following parameterizations of the physical processes were used: cloud microphysics [9]; long-wave radiation [10]; short-wave radiation [11]; boundary layer [12]; surface layer [13]; and soil layer [14]. For calculations involving the urban canopy model UCM [15], the following parameterizations were used: cloud microphysics [16] and boundary layer [17]. The WRF versions 3.5.1 and 3.8.1 were used.

The initial and boundary conditions for the fields of meteorological values were specified using data of the FNL model (NCEP, global analysis data) with a time step of 6 hours and a horizontal resolution of 0.5 degrees to calculate the first "coarse" domain D1. The initial and boundary conditions for the fields of chemical substances were set to zero.

3. Urban parameters

The sensitivity of the WRF modeling results and the IMDAF inverse problem results to the influence of inclusion of the single-level urban canopy model (SLUCM), as well as to a change of some of its
parameters and a priori data was studied. The SLUCM describes the processes of thermal and radiation balance by using a priori information of the urban area classification, building height, building materials (for roofs and walls) and their empirical characteristics (roughness, heat capacity, reflectivity, and emissivity). A significant parameter is the percentage of artificial structures and natural vegetation for each model cell [3]. Thus, simulation of the effect of urban agglomeration on the meteorological characteristics, as well as on the gas and aerosol distribution, requires obtaining some static parameters for the whole domain or for each model cell. For this, the georeferenced data of the OpenStreetMap project (OSM) [18] were processed.

The categories of urban land use [15] and non-urban categories have been updated (Figure 1) using the IGBP classification [19]. The categories were updated into the WRF initial data file. Analysis of intersection of geo-objects with the cell geometry was made for each model cell. The fraction of non-urban categories was taken as the area of intersection of the object category geometry and the model cell. The total area of artificial constructions presented the urban fraction (Figure 2), and the total area of natural objects was given for the vegetation fraction.

The OSM includes a large number of classified land use objects in accordance with the CORINE land cover classes for the Novosibirsk agglomeration region. A dictionary of correspondence of OSM objects to the WRF table land cover classes [20] was created.

![Figure 1. Land use categories obtained by processing OSM data for domain D2.](image1)

![Figure 2. Urban fraction obtained by processing OSM data for domain D2.](image2)

The parameters of the three urban categories were obtained using OSM data: the size and location of buildings, their deviations from the average. The albedo values were corrected based on measurement data [21] in accordance with the season of the model experiment. The roughness length values for each of the land classes were updated in accordance with the recommendations [22]. The roughness parameter value was set at 0.75 meters for the category of artificial structures.

4. Emission source reconstruction method

4.1. A priori transport traffic information

The IMDAF model uses information on the location of emission sources and their temporal characteristics. All sources for the experiments were divided into sources with constant intensity (CHP, boiler house emission rates are considered constant for simplicity) and sources with variable intensity (transport). All sources are considered as point sources. The grid of the D2 WRF domain coincides with the horizontal IMDAF domain. The location of CHP and boiler plant sources (triangular markers in Figure 3) was selected as the nearest node to the coordinates of the object. The archive of information accumulated from Yandex Static API [23] was used to obtain the location of transport sources on the grid D2 WRF domain (green square symbols in Figure 4) and their temporal
characteristics. The data provided by the service is a raster image generated from georeferenced data and processed traffic information on the road sections.

The average values of the relative number of pixels of road traffic were obtained from the series of half-hour data from the Yandex Static API for 2015-2018. The values corresponding to Sunday, Monday, Tuesday, and Wednesday of each week of July were used for this traffic intensity factor. The resulting coefficient was used to solve the inverse source problem with the IMDAF.

4.2. Inverse source problem solution algorithm

Let there be a uniform temporal grid with the step size $\Delta t$ and spatial grids:

$$
\omega_x = \left\{ \{ x_j \}_{j=1}^N \mid 0 = x_1 < \ldots < x_j < \ldots < x_N = X \right\},
$$

$$
\omega_y = \left\{ \{ y_j \}_{j=1}^N \mid 0 = y_1 < \ldots < y_j < \ldots < y_N = Y \right\}.
$$

On the grid there is the discretized model

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

$$
\frac{\phi^k - \phi^{k-1}}{\Delta t} = L \phi^k + F^k + \psi^k, \quad k = 2, \ldots, N_t,
$$

where $\phi^k$ is the state function, $\varphi_0$ is the initial condition, $F$ is the aggregate of the boundary conditions, and $f$ is the source function. Let the solution of the direct problem be the solution of (1), (2) with given initial $\varphi_0$ and boundary conditions and the source $f$. In the inverse problem, $f$ is unknown, instead there are $M$ concentration measurements $\mathcal{T} = \{ \{ t_m \}_{m=1}^M \}$ at the grid points $\{(x_m, \theta_m)\}_{m=1}^M \subset \omega_x \times \omega_y$:

$$
\varphi(x_m, \theta_m) = I_m + \delta I_m.
$$

By means of the adjoint problems, we can connect the state function to the source $f$:

$$
\sum_{k=2}^N \langle f^k, \psi^k \rangle \Delta t = \sum_{k=2}^N \langle \phi^k [f] - \phi^k [0], h^k \rangle,
$$

where $\psi^k$ is the solution of the adjoint problem

$$
\psi^k = 0, \quad k = N_t + 1,
$$

$$
\frac{\psi^k - \psi^{k+1}}{\Delta t} = (L^*)^k \psi^k + h^k, \quad k = 2, \ldots, N_t,
$$

and $^*$ denotes the operation of taking the adjoint with respect to the scalar product

$$
\langle a, b \rangle = \sum_{i,j} a_{ij} b_{ij} \delta x_i \delta y_j, \quad \delta x_i = \begin{cases} \Delta x / 2, & i = 1 \\ \Delta x, & 1 < i < N \end{cases}, \quad \delta y_j = \begin{cases} \Delta y / 2, & j = 1 \\ \Delta y, & 1 < j < N \end{cases}.
$$

Let us denote the adjoint problem solution by $\psi[h]$. With the help of the scalar product we can rewrite the measurement data in the form

$$
\sum_{k=2}^N \langle \phi^k, \delta(i - \tilde{i}_m, j - \tilde{j}_m, k - \tilde{k}_m) \rangle = \phi^k_{w,m} = I_m, \quad m = 1, \ldots, M, \quad \delta(i - \tilde{i}, j - \tilde{j}, k - \tilde{k}) = \frac{\delta x_i \delta y_j}{\delta x_i \delta y_j} \delta_{ii} \delta_{jj} \delta_{kk}.
$$
where \( \left\{ \left( \tilde{i}_m, \tilde{j}_m, \tilde{k}_m \right) \right\}_{m=1}^M \) are the indices of the grid points with the coordinates \( \left\{ \left( \chi_m, \theta_m \right) \right\}_{m=1}^M \). Let

\[
\psi_m := \psi[\delta(\cdots, -\tilde{i}_m, \cdots, -\tilde{j}_m, \cdots)], \quad m = 1, \ldots, M.
\]

With these adjoint functions, we have the system of \( M \) equations

\[
\sum_{k=2}^{N} \left\{ f_k, \psi_m^k \right\} \Delta t = I_m - \sum_{k=2}^{N} \left\{ \phi_k, \delta(\cdots, -\tilde{i}_m, j - \tilde{j}_m, k - \tilde{k}_m) \right\} = \bar{T}_m, \quad m = 1, \ldots, M. \tag{4}
\]

To incorporate the \textit{a priori} information on the sources in the inverse problem solution, we use the following parameterization of \( f \) :

\[
f^k_{ij} = \sum_{s=1}^{S} q_{s} \delta(i - \tilde{i}_s, j - \tilde{j}_s) + \sum_{r=1}^{R} q_{r} Q(t^r) \delta(i - \tilde{i}_r, j - \tilde{j}_r), \quad \delta(i - \tilde{T}, j - \tilde{J}) = \frac{\delta_{x} q_{x} \delta_{y} q_{y}}{\delta x_{\tilde{Y}}, \delta y_{\tilde{Y}}}, \quad \delta_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}
\]

\[\text{Figure 3. Emission distribution: green squares are transport emission points, triangles are CHPs, rhombuses are observation stations.}\]

\[\text{Figure 4. Recovered power of emissions for preset time. Grey fields denote EDGAR scenario; red, orange, and green markers are Sol-1, Sol-2, and Sol-3 scenarios, respectively. The centers of markers correspond to point source locations and the radius of a marker is proportional to the source power.}\]

If we put the parameterization of \( f \) with the unknown \( \left\{ q_{s} \right\}_{s=1}^{S} \) into (4), we obtain the matrix equation

\[
S\bar{q} = \bar{T}, \quad S_{ss} = \begin{cases} 
\sum_{k=2}^{N} \left\{ \psi_m^k, \delta(\cdots, -\tilde{i}_m, \cdots, -\tilde{j}_m) \right\} \Delta t, \quad 1 \leq s \leq N \\
\sum_{k=2}^{N} Q(t^k) \left\{ \psi_m^k, \delta(\cdots, -\tilde{i}_m, \cdots, -\tilde{j}_m) \right\} \Delta t, \quad N + 1 \leq s \leq N + \bar{N}
\end{cases}
\]

To obtain a generalized solution we use the misfit cost functional and its gradient:

\[
J(\bar{q}) = \| S\bar{q} - \bar{T} \|^2, \quad \nabla J(\bar{q}) = 2S^T (S\bar{q} - \bar{T}).
\]
The exponential parameterization and the corresponding cost functional are used to obtain the nonnegative solution:

\[ q_s = e^{\beta_s}, \quad \beta_s \in \mathbb{R}, \quad s=1, \ldots, N + \bar{N}, \quad e^\beta = \left[ e^{\beta_1}, \ldots, e^{\beta_{N+\bar{N}}} \right]^T, \]

\[ J^*(\beta) = J(e^\beta), \quad \nabla J^*(\beta) = \text{diag}(e^\beta) \nabla J(e^\beta). \]

If \( \bar{\beta}^* \) is the minimum point of \( J^*(\beta) \), we consider \( q_s = e^{\beta_s}, \quad s=1, \ldots, N + \bar{N} \) as the solution of the inverse source problem.

### 4.3. Model validation

The results of the SO2 forecast were initially compared on a test set of emissions, since two different models, the WRF and the IMDAF 2D direct problem, were used in the numerical experiments. The inverse problem solution algorithm requires multiple solutions of the adjoint equations for the transport model. To make the source reconstruction algorithm computationally feasible, we use a simplified 2D model for the inverse problem solution. The question is how the results of the IMDAF 2D simulation could be compared with the results of the 3D WRF model. Table 1 shows the relative norm deviation between the solutions obtained at different model time intervals from the calculation start:

\[ e_{\text{IMDAF vs WRF}} = \left\| \phi_{\text{IMDAF}} - \phi_{\text{WRF}} \right\|_{L_2(\omega_x \times \omega_y \times \omega_z)} / \left\| \phi_{\text{IMDAF}} \right\|_{L_2(\omega_x \times \omega_y \times \omega_z)}, \]

where \( \phi_{\text{IMDAF}} \) is the IMDAF model solution, and \( \phi_{\text{WRF}} \) is the WRF-Chem model solution corresponding to the rows of Table 1. In the first row, the \( \phi_{\text{WRF}} \) fields were vertically averaged from the 0th to 7th levels and normalized to the mean height of 0-7 layers. Table 1 also shows the WRF-Chem solution at the 0th level, and the WRF-Chem solution at the 1st level. An analysis of the obtained prognostic concentrations shows that the IMDAF model qualitatively reproduces the time behavior of the averaged SO2 concentration obtained by the more complex WRF-Chem model. Probably the IMDAF has the best match when compared with the WRF data averaged over the vertical levels, because the same averaging procedure was used to set parameters from the WRF to the IMDAF. Note that the concentration maxima in the IMDAF model occur earlier than in the WRF-Chem model.

**Table 1.** Time-averaged normalized difference between solutions of IMDAF direct problem and WRF-Chem model

| Forecast interval, h | 0-6 | 06-12 | 12-18 | 18-24 | 24-30 | 30-36 | 36-42 | 42-48 | 48-54 | 54-60 | 60-66 | 66-72 | 0-72 |
|---------------------|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| WRF 0-7 level       | 0.34| 0.38  | 0.26  | 0.32  | 0.69  | 0.29  | 0.42  | 0.48  | 0.39  | 0.21  | 0.56  | 0.42  | 0.39  |
| WRF 0 level         | 0.41| 0.32  | 0.90  | 1.83  | 0.03  | 0.40  | 0.21  | 0.44  | 0.38  | 0.47  | 2.75  | 2.23  | 0.86  |
| WRF 1 level         | 0.35| 0.31  | 0.33  | 0.72  | 2.14  | 0.37  | 0.22  | 0.12  | 0.36  | 0.24  | 0.81  | 1.32  | 0.47  |

### 4.4. Emission reconstruction scheme

The horizontal wind components and the diffusion coefficients preliminarily calculated by the WRF model are used to solve the IMDAF inverse problem. The *a priori* information prepared for the IMDAF solutions is described in Section 4.1. The calculations were made from 00:00 UTC on July 13 to 00:00 UTC on July 17, 2008 to prepare the coefficient files and model concentration fields.
The IMDAF inverse problem solution for the case of given *a priori* information and measurement data was used for preparing the emission sets on the WRF D2 domain grid. The WRF scenario based on these emission data was compared with the scenario based on the EDGAR emissions. The results of the simulations were compared with the measurement data for the assessment of the quality of these emission sets.

Two ways of defining emission data were used. The initial definition was made by using the PREPCHM utility and the EDGAR global database v4.2 for 2008 [25]. In the first case, unchanged EDGAR emissions were used in the calculations of the WRF-Chem (to compare the simulation results of the developed composite system (Figure 5) with the results of the WRF-Chem modeling based on standard EDGAR emissions). In the second case, the EDGAR emissions in the resulting files were replaced with new values according to the emissions reconstructed by the IMDAF inverse problem solution. To harmonize the dimension of the emission sources used in the WRF-Chem and IMDAF models, an assumption on the regular distribution of the emission source power over the grid cell and its constancy in time for one hour was made. Thus, the first value from the time-dependent IMDAF values during each 1-hour period was selected. The emission rate of the IMDAF model has a dimension of kg/(m³·sec) and, therefore, the coefficient \( c = 10^6 \cdot 3600 / M \) was used for the transition to WRF-Chem emissions, where \( M = 64.066 \) g / moll is the molecular mass of SO2.

### 5. Numerical experiments for different WRF scenarios

A numerical experiment was performed for the period from 00:00 UTC on July 13 to 00:00 UTC on July 16, 2008. The dates were used because a relatively well-defined database of measurements could be obtained for this period. The first model day was used to adjust the model, and the subsequent ones were used to obtain the estimates.

The influence of urban parameterization on simulating the surface air temperature at the 2-meter level was studied. For the analysis, the observations of synoptic stations entering the domain D2 were used. The inclusion of the urban parameterization in the WRF version 3.5.1 corrected the daily maximums opposed to the calculation without the urban specificity. However, the night temperatures were significantly underestimated for the station inside the urban area, both as compared to the observations and to the calculation without urban parameterization. Also, the simulation of the WRF version 3.8.1 with the same parameters for the urban model was made. In this version, the influence of latent anthropogenic heat fluxes was taken into account (the default values were used). The result has shown more accurate values for the daytime maximum relative to the observational data in comparison with the calculation without urban parameterization. The night temperatures were similar to non-urban ones and were underestimated by an average of 2 degrees. The night temperatures modeled by the WRF-Urban 3.8.1 did not have an additional error as in the simulations with the WRF-Urban 3.5.1.

The measurement data from 00:00 UTC on July 13 to 00:00 UTC on July 17, 2008 collected at 5 stations (marked by rhombuses in Figures 3 and 4) of the Center for Environmental Monitoring of the...
West Siberian Hydrometeorological Service were used for estimations in the emission rate recovery scheme. The measured substances are dust, sulfur dioxide, carbon monoxide, nitrogen dioxide, nitrogen oxide, benzopyrene, hydrogen sulphide, phenol, soot, hydrogen fluoride, ammonia, formaldehyde, and metals in the atmosphere. Measurements of the SO2 concentration for each station (the local time measurement mode was 07:00, 13:00, and 19:00) were used in the present study.

The numerical experiment included four scenarios for simulating the SO2 distribution. In three scenarios, emission data were obtained from the IMDAF inverse problem. To prepare the IMDAF coefficients, the WRF simulations were carried out with various combinations of default and implemented land use categories, as well as different physical parameterizations.

The list of scenarios is as follows:

- EDGAR: emissions from the global EDGAR database, USGS land use categories, and parameterizations [9][12] were used;
- Sol-1: recovery of emissions by the IMDAF solution, USGS land use categories, and parameterizations [9][12] were used;
- Sol-2: recovery of emissions by the IMDAF solution, land use categories from OSM data and parameterizations [16][17] were used, urban parameterization was turned off; and
- Sol-3: recovery of emissions by the IMDAF, land use categories from OSM data and parameterizations [16][17] were used, urban parameterization was turned on.

Figure 4 shows the EDGAR emissions (grey square areas). The source power in the case of EDGAR is indicated by the shading intensity for each model cell that falls within these areas. Red and yellow circles and green hexagons correspond to the power recoveries of Sol-1, Sol-2, and Sol-3, respectively. The centers of the markers correspond to the point source locations, and the radius of the marker is proportional to the source power. In terms of the total emission power for the whole domain D2 per year, the values are: EDGAR: 0.68 KT, Sol-1: 2.62 KT, Sol-2: 1.76 KT, and Sol-3: 2.38 KT. According to [26], the total power of SO2 emissions for Novosibirsk in 2008 was 42.6 KT.

**Table 2. Comparison of simulation results with measurement data at observation station grid nodes**

| Station index | EDGAR | Sol-1 | Sol-2 | Sol-3 | Averaged measurements |
|---------------|-------|-------|-------|-------|-----------------------|
|               | WRF   | IMDAF | WRF   | IMDAF | WRF   | IMDAF | WRF   | IMDAF | µg/m³   |
| 1             | 0.08  | -0.05 | 0.11  | 0.43  | 0.34  | 0.52  | 0.24  | 0.57  |          |
| 19            | -1.70 | -1.41 | -0.85 | -0.23 | 0.17  | 1.00  | 0.24  | 2.14  |          |
| 26            | -0.63 | -0.34 | -0.48 | -0.27 | -0.39 | -0.14 | -0.49 | 1.14  |          |
| 49            | -4.00 | -3.49 | -1.91 | -3.27 | -2.17 | -2.77 | -1.96 | 4.29  |          |
| 54            | -0.64 | 0.01  | -0.70 | -0.40 | -0.82 | -0.20 | -0.59 | 1.00  |          |

Table 2 lists the results of comparison of the SO2 values calculated by the WRF-Chem and by the direct problem of the IMDAF for the four scenarios. The same emissions for both the WRF-Chem and IMDAF direct problem simulations are used. The average bias between the measured and calculated SO2 values at particular measurement time points is given for the coordinates of stations at grid nodes of the domain D2 (IMDAF) and spatially-averaged by the 25 closest grid points (WRF). The spatial averaging was not performed for the IMDAF, since the IMDAF solutions are fitted to the measurements. The comparison was performed for the SO2 values averaged over 0 to 7 vertical WRF levels and those of the direct IMDAF problem. We use the averaged vertical SO2 concentrations in the WRF estimates, since similar averaging of the meteorological quantities is used in the emission reconstruction scheme and in solving the direct IMDAF problem.

An analysis of the Table shows significantly underestimated concentrations for the EDGAR and Sol-1 scenarios for station indexes 19, 26, and 49 where the average values are at least 1 µg / m³ of
SO2 in a three-day period. It can be explained by wrongly estimated emission intensity (Sol-1) and wrong location of the sources for these scenarios. Figure 4 demonstrates that the EDGAR has a larger number of emission grid points than the other scenarios, but with comparable total intensity. In the WRF model scenarios from Sol-1 to Sol-3, a trend to rising of the SO2 level at the observation points is obtained. The WRF solutions in the Sol-3 scenario are most realistic for forecasting air pollution close to the grid cells which contain the observation stations. Significant differences between the values of the scenario results and observations are due to a rare observation network both by number of stations and measurements per day. The mismatches between the results of the IMDAF and WRF may be due to the time shifts of the IMDAF concentration maxima relative to those of the WRF.

6. Conclusions

An offline coupled forecasting system consisting of a chemical transport model, IMDAF, and the open source WRF model with a Chem component was implemented with the purpose to improve air quality forecasts in the absence of operational information on emission sources. The system includes calculations and preparation of the meteorological elements produced by the WRF, data of an air quality monitoring system and a priori information about the location and the temporal behaviour of emission sources. These data are used by the IMDAF to assign the source parameters for the chemical transport model. The solution of the inverse IMDAF problem is utilized to build emission sets in the WRF simulations of the pollutant concentration in a passive tracer mode. Various physical parameterizations, including additional urban data in the WRF simulations, were tested in the full cycle of the forecasting system. One iteration of the assimilation cycle of the city air pollution monitoring data with a three-day assimilation window was presented.

In the above scenarios, where the WRF Single-Moment 6-Class cloud schemes and Mellor, Jamada, Yanyic boundary layer parameterization schemes were used, the recovered emission intensity is more plausibly distributed over the CHPs sources. Using an urban parameterization in the WRF simulations reduces the simulation underestimation at the grid points with observations.

Due to the complexity of the inventory of all Novosibirsk agglomeration emission sources, the use of a larger number of measurement stations, mobile monitoring, and an inventory of the location and type of the sources can improve the quality of the forecast.

Acknowledgments

The adaptation of the direct and adjoint models to city conditions, as well as the numerical experiments, were carried out with support of the RFBR and the Novosibirsk Regional Government (project no. 17-41-543309).

References

[1] WHO Regional Office for Europe 2006 Air Quality Guidelines Global Update 2005: Particulate Matter, ozone, nitrogen dioxide and sulfur dioxide (A EURO Publication) World Health Organization

[2] Penenko V V, Penenko A V, Tsvetova E A 2017 Variational approach to the study of processes of geophysical hydro-thermodynamics with assimilation of observation data Journal of Applied Mechanics and Technical Physics 58 771-778

[3] Chen F, H Kusaka, R Bornstein, J K Ching, C Grimmond, S Grossman-Clarke, T Loridan, K W Manning, A Martilli, S Miao, D Sailor, F P Salamanca, H Taha, M Tewari, X Wang, A A Wyszogrodzki, C Zhang 2011 The Integrated WRF/urban modeling system: development, evaluation, and applications to urban environmental problems Int. J. Climatol. 31 273-288

[4] Fisher B, Chemel C, S Sokhi R, Francis X, Vincent K, J Dore A, Griffiths S, Sutton P, D Wright R 2015 Regional air quality models and the regulation of atmospheric emissions Quarterly Journal of the Hungarian Meteorological Service 119 355-378

[5] Markakis K, Valari M, Perrussel O, Sanchez O, Honore C 2015 Climate-forced air-quality modeling at the urban scale: sensitivity to model resolution, emissions and meteorology
Atmospheric Chemistry and Physics 15 7703-7723

[6] Holnicki P, Nahorski Z 2015 Emission Data Uncertainty in Urban Air Quality Modeling—Case Study Environmental Modeling & Assessment 20 583-597

[7] Penenko A V 2008 Identification of pollutant sources using variational technique Computational technologies 13 44-50

[8] Grell G A, Peckham S E, Schmitz R, McKeen S A, Frost G, Skamarock W C, Eder B 2005 Fully coupled online chemistry within the WRF model Atmospheric Environment 39 6957-75

[9] Hong S Y , Dudhia J, Chen S H 2004 A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation Monthly Weather Review 132 103-120

[10] Iacono M J, Delamere J S, Mlawer E J, Shephard M W, Clough S A, Collins W D 2008 Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models Journal of Geophysical Research 113

[11] Dudhia J 1989 Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model Journal of the Atmospheric Sciences 46 3077-3107

[12] Hong S Y, Noh Y, Dudhia J A 2006 New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes Monthly Weather Review 134 2318-41

[13] Beljaars A C M 1995 The parametrization of surface fluxes in large-scale models under free convection Quarterly Journal of the Royal Meteorological Society 121 255-270

[14] Chen F, Dudhia J 2001 Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity Monthly Weather Review 129 569-585

[15] Kusaka H, F Kimura 2004 Coupling a single-layer urban canopy model with a simple atmospheric model: Impact on urban heat island simulator for an idealized case J. Appl. Meteor. 43 1899-1910

[16] Hong S Y, Lim J-O J 2006 The WRF Single-Moment 6-Class Microphysics Scheme (WSM6) Journal Of The Korean Meteorological Society 42 129-151

[17] Mellor G L, T Yamada 1982 Development of a turbulence closure model for geophysical fluid problems Rev. Geophys. 20 851–875

[18] OpenStreetMap Wiki, 2018 http://wiki.openstreetmap.org

[19] Loveland T R, Reed B C, Brown J F, Ohlen D O, Zhu J, Yang L, and Merchant J W 2000 Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data International Journal of Remote Sensing 21, no. 6/7 1303-30

[20] T E Samsonov, P I Konstantinov 2014 OpenStreetMap data assessment for extraction of urban land cover and geometry parameters required by urban climate modeling Extended Abstracts of the GIScience 2014 40 395-399

[21] B D Belan, T K Skylyadneva, N V Uzhegova 2005 Razlichiyai albedo podstilayushey poverhnosti g. Novosibirska i ego okrestnostey Optika atmosferi i okeana 18 no. 3 238-241

[22] ECMWF 2014 Physical Processes IFS Documentation, Part IV https://www.ecmwf.int/sites/default/files/elibrary/2017/17736-part-iv-physical-processes.pdf

[23] Yandex Static API 2018 https://tech.yandex.ru/maps/doc/staticapi/1.x/

[24] Issartel J-P 2005 Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation Atmospheric Chemistry and Physics, Copernicus GmbH 5 249-273

[25] EC-JRC/PBL 2011 Global Emissions EDGAR v4.2 http://edgar.jrc.ec.europa.eu/overview.php?v=42

[26] E V Bezuglaya 2009 Ezhegodnik sostoyanie zagryazneniya atmosfery v gorodakh na territorii Rossii za 2009 g (St. Petersburg: GGO im. A.I.Voejkova) p 84