

Article

# Sensitivity Operator Framework for Analyzing Heterogeneous Air Quality Monitoring Systems

Alexey Penenko <sup>1,2,\*</sup>, Vladimir Penenko <sup>1,2</sup>, Elena Tsvetova <sup>1</sup>, Alexander Gochakov <sup>1,3</sup>, Elza Pyanova <sup>1</sup>  
and Viktoriia Konopleva <sup>1,2</sup>

<sup>1</sup> Institute of Computational Mathematics and Mathematical Geophysics SB RAS, 630090 Novosibirsk, Russia; penenko@sscc.ru (V.P.); e.tsvetova@ommgp.sccc.ru (E.T.); gochakov@sibnigmi.ru (A.G.); pyanova@ommgp.sccc.ru (E.P.); v.konopleva@g.nsu.ru (V.K.)

<sup>2</sup> Department of Mathematics and Mechanics, Novosibirsk State University, 630090 Novosibirsk, Russia

<sup>3</sup> Siberian Regional Hydrometeorological Research Institute, 630099 Novosibirsk, Russia

\* Correspondence: aleks@ommgp.sccc.ru; Tel.: +7-383-330-6152

**Abstract:** Air quality monitoring systems differ in composition and accuracy of observations and their temporal and spatial coverage. A monitoring system's performance can be assessed by evaluating the accuracy of the emission sources identified by its data. In the considered inverse modeling approach, a source identification problem is transformed to a quasi-linear operator equation with the sensitivity operator. The sensitivity operator is composed of the sensitivity functions evaluated on the adjoint ensemble members. The members correspond to the measurement data element aggregates. Such ensemble construction allows working in a unified way with heterogeneous measurement data in a single-operator equation. The quasi-linear structure of the resulting operator equation allows both solving and predicting solutions of the inverse problem. Numerical experiments for the Baikal region scenario were carried out to compare different types of inverse problem solution accuracy estimates. In the considered scenario, the projection to the orthogonal complement of the sensitivity operator's kernel allowed predicting the source identification results with the best accuracy compared to the other estimate types. Our contribution is the development and testing of a sensitivity-operator-based set of tools for analyzing heterogeneous air quality monitoring systems. We propose them for assessing and optimizing observational systems and experiments.

**Keywords:** air quality; monitoring systems; transport and transformation of impurities; inverse problem; emission source identification; sensitivity operator; adjoint equations; Lake Baikal region



**Citation:** Penenko, A.; Penenko, V.; Tsvetova, E.; Gochakov, A.; Pyanova, E.; Konopleva, V. Sensitivity Operator Framework for Analyzing Heterogeneous Air Quality Monitoring Systems. *Atmosphere* **2021**, *12*, 1697. <https://doi.org/10.3390/atmos12121697>

Academic Editors: Gennadii Matvienko and Oleg Romanovskii

Received: 14 November 2021

Accepted: 13 December 2021

Published: 18 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Atmospheric air quality monitoring systems vary in their temporal and spatial coverage, the composition of the chemicals observed, and the accuracy of the data obtained. With the help of various monitoring systems, a huge amount of heterogeneous data is collected. The effective use of data is a significant scientific problem [1,2]. Moreover, new measurement systems are deployed to obtain new air-quality data, and new observational experiments are carried out. Choosing or optimizing the configuration of a monitoring system or observational experiment [3–11] is a multi-criteria decision-making task. It is necessary to compare the cost and accuracy of measuring equipment, maintenance costs, the goals of building such a system, its use cases, and other factors [12–14]. An essential criterion for choosing the configuration of the monitoring system is the value of the information content of prospective measurements. Hence, the methods of different accuracy and computational intensity for analyzing the information content of measurement data are essential. Finally, all the factors have to be compared, and a decision should be made about choosing the best variant. To choose one of the options, one can, for example, use the Multiple Criteria Decision-Making methods (see, e.g., [15]).

The joint use of process models and observational data makes it possible to solve a number of applied inverse problems, including the identification of air pollution sources and restoration of pollution fields in unobserved areas [3,4,16–23]. Data assimilation algorithms [24,25] are widely used in operative and quasi-operative air quality applications [26–31]. A data assimilation problem can be considered as the sequence of linked sources identification problems [32]. Air pollution sources pattern extracting is needed to compile emission inventories [33]. Knowing the emission sources is essential to model air quality since the emission sources largely determine the system's behavior [34,35]. These considerations make the source identification problem's solution accuracy a relevant instrument to measure the value of the information content of the observation data.

The methods of utilizing heterogeneous measurement data are referred to as “data fusion” [36–38]. Data of different sources and types can complement and cross-validate each other. In the paper, we focus on data fusion of heterogeneous measurements from the perspective of inverse modeling. In particular, previously, we considered the data obtained from the time series of concentrations at monitoring sites [39,40] and those were presented as images of concentration fields [41,42], as well as a combination of these two data types [43]. In the paper, we additionally consider Pointwise (in situ) monitoring data in space and time, which allows processing data from mobile monitoring systems mounted on ships [44], aircraft [45], unmanned aerial vehicles [46], and cars [47]. Moreover, the support for time-averaged concentration data is incorporated that allows one to include, for example, snow cover monitoring data, bio-monitoring techniques data, as well as the other, averaged data, for example, those published by Roshydromet [48].

Our approach is based on ensembles of solutions of adjoint equations and sensitivity operators. It consists of constructing families of sensitivity operators of various dimensions based on the inverse problems, linking the observed characteristics with the model's uncertainty functions (e.g., emission sources). The approach bridges an idea proposed by G.I. Marchuk in [49], methods of sensitivity theory [50], the essence of the mollification method [51], and the “image to structure” operator concept [52]. In [49], G.I. Marchuk suggested gathering adjoint problem-based interpretations of separate measurement data elements to a single equation. In the case of many data elements (e.g., high-resolution satellite images), the straightforward application of the approach is complicated since one needs to solve a large number of adjoint problems, equal to the number of the considered data elements (e.g., pixels).

The first approach to this issue is considering a misfit cost function comparing the measured and simulated data. It dramatically decreases the ensemble of the adjoint problems to obtain the gradient of the cost function [53–55]. In this case, one has to solve only one adjoint problem, defined by the misfit between the measured and simulated data. This approach is widely used in the applications (see the review in [24,25]).

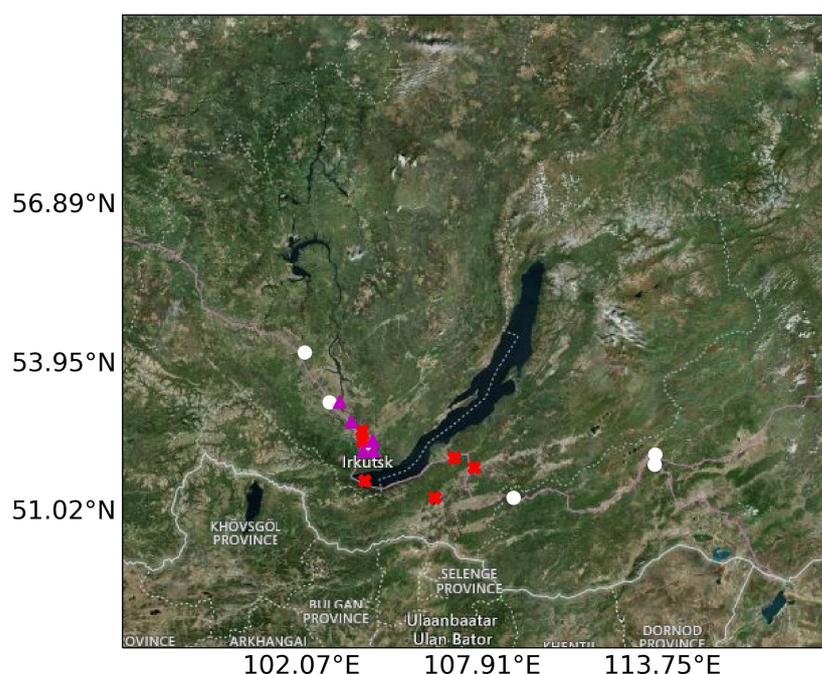
Another solution is to consider some arbitrary aggregates of the measurement data (which is relevant to the mollification method and the “image to structure” operator concept) and evaluate the set of adjoint problem solutions for these aggregates. The sensitivity operator is then composed of this adjoint problem solutions ensemble. Sensitivity operators allow formulating a family of quasi-linear operator equations for the inverse problem, the sets of solutions of which contain the solutions of the original problem. The quasi-linear structure of the equations provides a more straightforward way of analysis with the linear tools [39,41], compared to the case of the misfit cost function-based aggregation of the measurement data, i.e., the sensitivity operators can be used to solve and analyze the inverse problems in a unified way. Similar algorithms have been applied to the linear passive transport problem [56] and the nonlinear transport–transformation model with Pointwise sources and in situ measurements [20]. An overview of the other adjoint-ensemble-based methods can be found in [23].

Furthermore, the sensitivity of the operator-based approach allows naturally fusing various data types [57]. To do this, it is enough to construct the sensitivity operator from the

adjoint problem solution ensembles corresponding to multiple data types (see Section 2.3 for more details).

The paper's objective is to briefly describe the features of the proposed approach to solving inverse modeling problems using heterogeneous data coming from various monitoring systems. In particular, we want to study the characteristics of the sensitivity operator that can be used to predict the source identification results based on different monitoring system configurations. The ultimate goal of the study is to find a few characteristics that can be used in the observation system optimization that includes various data sources. The purpose of such optimization in the context of source identification is to reduce the source identification error by configuring the observation network. We can solve the corresponding inverse problems for different observation system configurations to measure this error, but their solution may take significant time. That is why we are looking for approximate estimates that are computationally cheaper.

The approach is tested on the air quality inverse modeling scenario for the domain containing the Baikal Natural Territory (BNT) (Figure 1). In 1996, UNESCO recognized Lake Baikal as a World Heritage Site [58]. In 1999, the Federal Law on the protection of Lake Baikal was adopted, which defined the concept of BNT, where environmental zoning was carried out, prescribing special conditions for implementing economic activities in each zone [59]. The Baikal region is located in Eastern Siberia in the belt of temperate latitudes, considerably far away from the oceans. However, the big water masses of about 23,000 km<sup>3</sup> of fresh lake's water leave their mark on the climate and weather in the region. The complex terrain, including high mountain ranges that frame the lake, contributes to the emergence of strong local winds, which even have unique names: Barguzin, Shelonik, Gornaya, etc. The region is exposed to various anthropogenic and natural factors. For example, it suffers from forest fires of different origins [60–63]. The industrial activity that develops in the region, mainly near urban settlements, certainly affects the quality of the environment [64–66].



**Figure 1.** The geographical domain with monitoring sites locations “measuring” time series (red crosses); Pointwise measurements in space and time (white circles); Integrals over a time interval (magenta triangles).

This unique natural object, BNT, is observed by various means [44,63,67–71]. Hence, adequate measurement data-fusion is essential to carry out multidisciplinary studies in the

region and provide the governing authorities with the necessary data to make decisions in terms of environmental protection to ensure the region’s sustainable development.

The paper is organized as follows: In Section 2.1, we state the basic source identification problem. Section 2.2 presents our sensitivity-operator-based framework. The supported measurement types are indicated in Section 2.3. Some approaches to analyze the measurement system data are considered in Section 2.4. Section 2.5 describes the inverse modeling scenario for BNT. In Section 3, we analyze the numerical results and discuss the conclusions in Section 4.

## 2. Materials and Methods

### 2.1. Chemical Transport Model

The Chemical Transport Model (CTM) for  $l = 1, \dots, N_c$  reacting substances is considered in a domain  $\Omega_T = \Omega \times (0, T)$ , where  $\Omega$  is a sufficiently smooth approximation of a bounded rectangular domain  $[0, X] \times [0, Y]$  in  $\mathbb{R}^2$ ,  $T > 0$ . The domain  $\Omega_T$  is bounded by  $\partial\Omega_T = \partial\Omega \times [0, T]$ .

$$\frac{\partial \varphi_l}{\partial t} - \nabla \cdot (\text{diag}(\mu_l) \nabla \varphi_l - \mathbf{u} \varphi_l) + P_l(t, \boldsymbol{\varphi}) \varphi_l = \Pi_l(t, \boldsymbol{\varphi}) + f_l + r_l, \quad (\mathbf{x}, t) \in \Omega_T, \quad (1)$$

$$\mathbf{n} \cdot (\text{diag}(\mu_l) \nabla \varphi_l) + \beta_l \varphi_l = \alpha_l, \quad (\mathbf{x}, t) \in \Gamma^{(out)} \subset \partial\Omega \times [0, T], \quad (2)$$

$$\varphi_l = \alpha_l, \quad (\mathbf{x}, t) \in \Gamma^{(in)} \subset \partial\Omega \times [0, T], \quad (3)$$

$$\varphi_l = \varphi_l^0, \quad \mathbf{x} \in \Omega, t = 0, \quad (4)$$

where  $t$  is time and  $\mathbf{x}$  is space coordinate,  $N_c$  is the number of considered substances,  $\varphi_l = \varphi_l(\mathbf{x}, t)$  denotes the concentration of the  $l^{th}$  substance at a point  $(\mathbf{x}, t) \in \Omega_T$ ,  $\boldsymbol{\varphi}$  is the vector of  $\varphi_l(\mathbf{x}, t)$  for  $l = 1, \dots, N_c$ , which is called the state function, and  $L = \{1, \dots, N_c\}$ . The functions  $\mu_l(\mathbf{x}, t) \in \mathbb{R}^2$  correspond to the diffusion coefficients,  $\text{diag}(\mathbf{a})$  is the diagonal matrix with the vector  $\mathbf{a}$  on the diagonal,  $\mathbf{u}(\mathbf{x}, t) \in \mathbb{R}^2$  is the underlying wind speed.  $\Gamma^{(in)}$  and  $\Gamma^{(out)}$  are parts of domain boundary  $\partial\Omega_T$  in which the vector  $\mathbf{u}(\mathbf{x}, t)$  points inwards the domain  $\Omega_T$  and is zero or points outwards the domain  $\Omega_T$ , respectively,  $\mathbf{n}$  is the outer normal. The functions  $\alpha_l(\mathbf{x}, t)$  and  $\varphi_l^0(\mathbf{x})$  describe the boundary and initial conditions, respectively,  $\beta_l$  is the boundary condition parameter,  $f_l(\mathbf{x}, t)$  is the a priori known source function. The transformation model defines loss and production operator elements  $P_l, \Pi_l : [0, T] \times \mathbb{R}_+^{N_c} \rightarrow \mathbb{R}_+$ . In the considered case of the chemical transformations, they are polynomials with positive time-dependent coefficients.

In our approach to inverse modeling, the model parameters are divided into “predefined parameters”  $\mathbf{v}$  and “uncertainty functions”  $\mathbf{q}$ , belonging to some set  $Q$ . Most often, the uncertainty functions in the air quality studies are unknown sources of pollution  $\mathbf{q} = \mathbf{r}$ . To connect the current work with the other papers of the series, we intentionally keep a more complicated notation than it is needed to describe the considered problem. The presented approach is general and can be used as a framework for a broad spectrum of different inverse modeling problems. Let  $\mathbf{r} \in Q$ , where  $Q$  is the set of admissible sources, defined by a priori information:

- We suppose that only a given set of species  $L_{src}$  is emitted. For the rest of species  $r_l(\mathbf{x}, t) = 0, l \notin L_{src}$ .
- The emission sources are supposed to be constant in time ( $r_l(\mathbf{x}, t) = r_l(\mathbf{x})$ ).
- We do not require the emission sources to be positive since variables unconsidered in the model, chemical transformations, various land types, and meteorological conditions, such as rains and snowfalls, can act as sinks for the specific chemicals.

The rest of the parameters are considered as predefined, hence  $\mathbf{v} = \{\mu, \mathbf{u}, \alpha, \beta, \varphi^0, f\}$ . Let us define the pair of “direct” and “inverse” problems:

- In the Direct problem,  $\mathbf{v}$  and  $\mathbf{q} \in Q$  are given, and we find  $\boldsymbol{\varphi}$  from (1)–(4). The solution of the direct problem is denoted by  $\boldsymbol{\varphi}[\mathbf{q}]$ . Let there be an “exact” uncertainty function

value  $\mathbf{q}^{(*)} = \mathbf{r}^{(*)}$  to be found and  $\boldsymbol{\varphi}^{(*)} = \boldsymbol{\varphi}[\mathbf{q}^{(*)}]$  be the corresponding solution of the direct problem with the source function  $\mathbf{r}^{(*)}$ .

- In the Inverse problem, the uncertainty function  $\mathbf{q}^{(*)}$  has to be identified from the partial information (“measurement data”) about  $\boldsymbol{\varphi}^{(*)}$ , described in Section 2.3.

### 2.2. Sensitivity-Operator Based Representation of Measurement Data

We use the algorithm based on the sensitivity operators described in [39,41] to solve the source identification problem. It is based on the sensitivity relation, which links the model state function variation with that of the uncertainty function. For any  $\mathbf{q}^{(2)}, \mathbf{q}^{(1)} \in Q$ :

$$\langle S[\mathbf{q}^{(2)}, \mathbf{q}^{(1)}; h], \mathbf{q}^{(2)} - \mathbf{q}^{(1)} \rangle_Q = \langle h, \delta\boldsymbol{\varphi} \rangle_H. \tag{5}$$

Here  $S[\mathbf{q}^{(2)}, \mathbf{q}^{(1)}; h]$  denotes the sensitivity function, which is calculated by the solution of the adjoint problem (the details can be found in [41]). The solution of the adjoint problem is determined by its source function  $h$ , which is defined by the measurement operator (see Section 2.3). Scalar products in (5) are:

$$\langle a, b \rangle_H = \sum_{l=1}^{N_c} \int_0^T \int_{\Omega} a_l(\mathbf{x}, t) b_l(\mathbf{x}, t) d\mathbf{x} dt, \quad \langle a, b \rangle_Q = \sum_{l=1}^{N_c} \int_{\Omega} a_l(\mathbf{x}) b_l(\mathbf{x}) d\mathbf{x}. \tag{6}$$

The right-hand side corresponds to an element of the measurement data.

If we consider a set of functions  $U = \{h^{(\xi)}\}_{\xi=1}^{\Xi}$ , then combining the corresponding relations (5), we obtain a sensitivity operator relation

$$M_U[\mathbf{q}^{(2)}, \mathbf{q}^{(1)}] (\mathbf{q}^{(2)} - \mathbf{q}^{(1)}) = H_U \boldsymbol{\varphi}[\mathbf{q}^{(2)}] - H_U \boldsymbol{\varphi}[\mathbf{q}^{(1)}], \tag{7}$$

where

$$M_U[\mathbf{q}^{(2)}, \mathbf{q}^{(1)}] z = \sum_{\xi=1}^{\Xi} \mathbf{e}^{(\xi)} \langle S[\mathbf{q}^{(2)}, \mathbf{q}^{(1)}; h^{(\xi)}], z \rangle_Q, \quad H_U \boldsymbol{\varphi} = \sum_{\xi=1}^{\Xi} \mathbf{e}^{(\xi)} \langle h^{(\xi)}, \boldsymbol{\varphi} \rangle_H.$$

Here  $\mathbf{e}^{(\xi)}$  is the  $\xi$ -th element of the canonical basis in  $\mathbb{R}^{\Xi}$ . The adjoint functions corresponding to different elements of  $U$ , needed to compose the sensitivity operator, can be evaluated in parallel as an ensemble. Combining the ensembles of solutions of adjoint equations corresponding to different measurement data types into one sensitivity operator allows us to consider heterogeneous data.

If  $\mathbf{q}^{(*)}$  is the exact solution of the source identification problem,  $I$  is the measurement data, aggregated in the state-function form (i.e., it is equal to  $\boldsymbol{\varphi}^{(*)}$  in the parts of  $\Omega_T$  where there are measurements and is zero otherwise), and  $\delta I$  is its perturbation (i.e., the measurement noise), then for any  $U$  and  $\mathbf{q}$  the relation holds:

$$M_U[\mathbf{q}^{(*)}, \mathbf{q}] (\mathbf{q}^{(*)} - \mathbf{q}) = H_U I + H_U \delta I - H_U \boldsymbol{\varphi}[\mathbf{q}]. \tag{8}$$

Quasi-linear operator Equation (8) can be solved by any appropriate operator equation solution method. In our work, we use the Newton–Kantorovich-type algorithm regularized with the truncated singular value decomposition (SVD) analogous to the algorithms in [41,42]. We present the algorithm in Appendix A. In this modification, we enforce the monotonic decrease of the discrepancy function with the iterations by choosing the appropriate step sizes.

### 2.3. Measurement Data Types

To use the approach, we have to present the measurement data in the form of a scalar product of the model state function with some projection functions. To define

the measurements and corresponding projection functions  $h$ , we need appropriate delta-functions over space, time, and species denoted by  $\delta$  and elements of the cosine-basis on the interval  $[0, X]$ :

$$C(X, \theta, x) = \frac{1}{\sqrt{X}} \begin{cases} \sqrt{2} \cos\left(\frac{\pi\theta x}{X}\right), & \theta > 0 \\ 1, & \theta = 0 \end{cases}.$$

The elements of cosine-basis are used to transform image-type data to a finite set of values through Fourier transform. The symbol  $(\cdot)_{meas}$  denotes the parts of the domain where there are measurements available. For example, usually, only a small set  $L_{meas} \subset \{1, \dots, N_c\}$  of chemical species is monitored.

We consider the following measurement types defined by their correspondence to the model state function and projection functions  $h^{(\xi)}$  in the sense of (6):

- “Timeseries”:  $N_{Timeseries}$  time series of concentrations of the specific species in the specific points. In the state function terms:

$$\left\{ \varphi_{l^{(m)}}(x^{(m)}, t), t \in [0, T], (x^{(m)}, l^{(m)}) \in (\Omega \times L)_{meas}, m = 1, \dots, N_{Timeseries} \right\}.$$

Projection system:

$$h^{(\xi)} = C(T, \theta^{(\xi)}, t) \delta(x - x^{(\xi)}) \delta(l - l^{(\xi)}), \quad \xi = 1, \dots, \Xi_{Timeseries}.$$

For any element of  $(x^{(m)}, l^{(m)})$ , the parameter  $\theta$  ranges from 0 to  $\Theta_{Timeseries} - 1$ . The number of the frequencies  $\Theta_{Timeseries}$  is the parameter of the projection system. This parameter is responsible for the temporal resolution of the considered data. Hence the total number of projection functions corresponding to the Timeseries is

$$\Xi_{Timeseries} = \Theta_{Timeseries} \times N_{Timeseries}.$$

- “Pointwise”:  $N_{Pointwise}$  Pointwise concentration measurements of the specific species at specific moments and specific points. In the state function terms:

$$\left\{ \varphi_{l^{(m)}}(x^{(m)}, t^{(m)}), (x^{(m)}, t^{(m)}, l^{(m)}) \in (\Omega_T \times L)_{meas}, m = 1, \dots, N_{Pointwise} \right\}.$$

Projection system:

$$h^{(\xi)} = \delta(x - x^{(\xi)}) \delta(t - t^{(\xi)}) \delta(l - l^{(\xi)}), \quad \xi = 1, \dots, \Xi_{Pointwise}.$$

The projection system is naturally defined by the measurement points. Hence the total number of the projection functions is  $\Xi_{Pointwise} = N_{Pointwise}$ . In the case of a large number of points, these data can be aggregated.

- “Integral”:  $N_{Integral}$  Integrals of concentrations over the time interval of the specific species in the specific points. In the state function terms:

$$\left\{ \int_0^T \varphi_{l^{(m)}}(x^{(m)}, t) dt, (x^{(m)}, l^{(m)}) \in (\Omega \times L)_{meas}, m = 1, \dots, N_{Integral} \right\}.$$

Projection system:

$$h^{(\xi)} = \delta(x - x^{(\xi)}) \delta(l - l^{(\xi)}), \quad \xi = 1, \dots, \Xi_{Integral}.$$

Here  $\Xi_{Integral} = N_{Integral}$ . Integral measurements are equivalent to “Timeseries” measurements with  $\theta^{(\xi)} = 0$ .

- “Snapshot”:  $N_{Snapshot}$  specific species concentration fields images at specific moments in time. In the state function terms:

$$\left\{ \varphi_{l^{(m)}}(x, t^{(m)}), x \in \Omega, (t^{(m)}, l^{(m)}) \in ([0, T] \times L)_{meas}, m = 1, \dots, N_{Snapshot} \right\}.$$

Projection system:

$$h^{(\xi)} = C(X, \theta_x^{(\xi)}, x)C(Y, \theta_y^{(\xi)}, y)\delta(t - t^{(\xi)})\delta(l - l^{(\xi)}), \quad \xi = 1, \dots, \Xi_{Snapshot}.$$

The projection system has two parameters:  $\Theta_{Snapshot}^{(x)}$  and  $\Theta_{Snapshot}^{(y)}$ , which define the spatial resolution of the considered data. For any image,  $\theta_x$  and  $\theta_y$  range in  $0, \dots, \Theta_{Snapshot}^{(x)} - 1$  and  $0, \dots, \Theta_{Snapshot}^{(y)} - 1$ , respectively. Hence  $\Xi_{Snapshot} = N_{Snapshot} \times \Theta_{Snapshot}^{(x)} \times \Theta_{Snapshot}^{(y)}$ .

In the numerical algorithm, the operator Equation (8) is transformed to the discrete form. Hence, the projection functions are also substituted by their discrete analogs. To unify the scales of the different projection functions [57], their discrete analogs are normalized with respect to a discrete version of scalar products (6). This normalization makes any element of measurement data equally important. To make some elements more important than others, they can be additionally weighted.

#### 2.4. Sensitivity-Operator-Based Analysis of Measurement System

Considering the quasi-linear representation (8) of the inverse problem data, we compare the following sensitivity-operator-based measurement system analyses.

##### 2.4.1. Inverse Problem Solution

The basic way to evaluate the informative content of the measurement data is to directly compare the result of solving (8) by a source identification algorithm with the known exact solution. The better the reconstruction, the more information is contained in the data. However, this estimation method requires a complete setting of the inverse modeling scenario (including the “exact” emission source), an appropriate algorithm of source identification, and enough time and computational capacities to solve the inverse problem. We call this approach “empirical”.

The procedure of evaluating the measurement system is carried out according to the following scheme:

1. The “exact” solution  $\mathbf{q}^{(*)}$  is given. In our case, this is the location and capacity of the emission sources.
2. The “exact” solution  $\mathbf{q}^{(*)}$  is then used to simulate the “measurement data”. This “measurement data” is used in the algorithm to solve the inverse problem.
3. The result  $\mathbf{q}^{(\infty)}$  of the algorithm is compared with the “exact” solution. In this case, both the reconstruction of the source is estimated, and the convergence parameters of the algorithm are analyzed.

This estimation strongly depends on the source identification algorithm, which may have its features. For example, it can be locally convergent, and therefore, dependent on the initial guess.

##### 2.4.2. Sensitivity Operator Properties Analysis

The quasi-linear inverse problem representation (8) can be used to predict the quality of solving the inverse problem prior to solving it [39,41]. Indeed, if we assume that  $M_U[\mathbf{q}^{(*)}, \mathbf{q}]$  is known or is weakly dependent on  $\mathbf{q}^{(*)}$ , then Equation (8) becomes a linear ill-posed operator equation, and we can use linear theory to evaluate the inverse problem properties.

Let there be grid domains  $\Omega^h$  and  $\Omega_T^h$ , corresponding to  $\Omega$  and  $\Omega_T$ , and a map that encodes a grid point with coordinates  $(x_i, y_j, l) \in \Omega^h \times L_{src}$  by a single index. Let  $N_x, N_y$

be the numbers of grid points of  $\Omega^h$  in the  $x$  and  $y$  dimensions, respectively. In the rest of the paper, we work with the discretized versions of the considered aggregates. Hence we can consider  $\mathbf{q}$  as the vector from  $\mathbb{R}^{N_x N_y |L_{src}|}$ . Here,  $|L_{src}|$  means the number of elements in  $L_{src}$ . The sensitivity operator  $M_U[\mathbf{q}^{(*)}, \mathbf{q}]$  can be considered as the matrix  $\mathbb{R}^{\Xi \times (N_x N_y |L_{src}|)}$ .

In the theory of ill-posed problems, the linearized version of the operator Equation (8) in the neighborhood of the exact solution:

$$M(\mathbf{q}^{(*)} - \mathbf{q}) \approx f_M, \quad f_M := H_U I - H_U \varphi[\mathbf{q}], \quad M := M_U[\mathbf{q}^{(*)}, \mathbf{q}^{(*)}]$$

is considered. According to p. 49 in [72], regularization algorithms search for the best approximate solution of  $\delta \mathbf{q}^\dagger$ , which in our case is defined as:

$$\delta \mathbf{q}^\dagger = (M^* M)^\dagger M^* f_M \approx (M^* M)^\dagger (M^* M)(\mathbf{q}^{(*)} - \mathbf{q}), \tag{9}$$

where  $(M^* M)^\dagger$  is the generalized Moore–Penrose inverse for  $M^* M$ . The aggregate  $M^* M$  is the Fisher information matrix. The aggregate  $(M^* M)^\dagger (M^* M)$  is an orthogonal projector on the orthogonal complement to the sensitivity operator’s kernel. Indeed, if the singular value decomposition of  $M$  is  $\{U, S, V\}$  and for any  $z$ :

$$Mz = \sum_{n=1}^N U_n s_n \langle V_n, z \rangle,$$

where  $N$  is the number of nonzero singular vectors. Then,

$$M^* Mz = \sum_{n=1}^N V_n s_n^2 \langle V_n, z \rangle, \quad (M^* M)^\dagger z = \sum_{n=1}^N V_n \frac{\langle V_n, z \rangle}{s_n^2},$$

and

$$(M^* M)^\dagger (M^* M)z = \sum_{n=1}^N V_n \langle V_n, z \rangle.$$

Roughly speaking, only the projection of  $\mathbf{q}^{(*)} - \mathbf{q}$  on the nonzero right singular vectors of the sensitivity operator matrix can be restored at best because the rest of the information about the uncertainty variation is lost.

The disadvantage of using  $(M^* M)^\dagger (M^* M)$  in the source identification problem is that the dimensionality of  $M^* M$ , which has to be pseudo-inverted, is the (number of unknowns  $\times$  number of unknowns), which is as large as  $N_x N_y |L_{src}|$ . Another way to express the projector is to use the equivalent (Appendix B) aggregate:

$$M^* (M M^*)^\dagger Mz = \sum_{n=1}^N V_n \langle V_n, z \rangle. \tag{10}$$

In this case, the matrix  $M M^*$  has the dimensionality (number of adjoint ensemble members  $\times$  number of adjoint ensemble members) or  $\Xi \times \Xi$ , which is under our control. The representation (10) is more useful when the number of adjoint ensemble members is less than the number of unknowns, and therefore, we use it in our calculations. Let:

$$Y[\mathbf{q}^{(2)}, \mathbf{q}^{(1)}] := M^* (M M^*)^\dagger M,$$

where  $M = M[\mathbf{q}^{(2)}, \mathbf{q}^{(1)}]$ . As the first predictor of the solution, we consider:

$$\mathbf{q}^{(p)} := \mathbf{q}^{(0)} + Y[\mathbf{q}^{(0)}, \mathbf{q}^{(0)}] (\mathbf{q}^{(*)} - \mathbf{q}^{(0)}). \tag{11}$$

Another estimate that seems to be even more appropriate is:

$$\mathbf{q}^{(e)} := \mathbf{q}^{(0)} + Y[\mathbf{q}^{(*)}, \mathbf{q}^{(0)}] (\mathbf{q}^{(*)} - \mathbf{q}^{(0)}). \tag{12}$$

Estimates (11) and (12) take the exact solution  $\mathbf{q}^{(*)}$ , and it takes less time to evaluate them than solving the inverse problem with an algorithm.

Let us consider the properties of  $Y = Y[\mathbf{q}^{(0)}, \mathbf{q}^{(0)}]$ , which are independent of the exact solution. In the ideal case,  $Y$  should be equal to the identity operator  $Id$ . It means that the sources can be potentially reconstructed exactly. If we apply this operator to a single source with a unitary emission rate located in the  $\kappa$ -th point, we obtain the result of its representation by the monitoring system. This single source corresponds to the vector  $[Id]_{\kappa}$  with only one unitary element on the  $\kappa$ -th place. In the inverse modeling scenario (see Section 2.5), we deal with the Pointwise sources and  $[Y]_{\kappa} = Y[Id]_{\kappa}$ . Hence, the  $\kappa$ -th column of  $Y$  describes how the unitary source, located in the  $\kappa$ -th point of the grid domain, is transformed by the inverse problem. In the ideal case,  $[Y]_{\kappa}$  should be equal to  $[Id]_{\kappa}$ .

We can characterize the difference between  $[Y]_{\kappa}$  and  $[Id]_{\kappa}$  on the Pointwise basis with several meaningful characteristics linked to the structure and interpretation of  $Y$ . The first characteristic is:

$$E_{\kappa}^{(n)} = \|[Id]_{\kappa} - [Y]_{\kappa}\|. \tag{13}$$

If the information about the source in the  $\kappa$ -th point is lost completely, then the norm  $\|[Y]_{\kappa}\|$  is zero. In the ideal case,  $\|[Y]_{\kappa}\| = 1$ . We consider the following error-type characteristic:

$$E_{\kappa}^{(e)} = 1 - \|[Y]_{\kappa}\|, \tag{14}$$

To evaluate the characteristics  $E_{\kappa}^{(n)}$  and  $E_{\kappa}^{(e)}$ , we have to evaluate the complete  $Y$  matrix. Let us look at the characteristics that are evaluated with the reduced amount of  $Y$ 's elements. In [19], there is an aggregate called "Illumination" and it is expressed as:

$$[M]_{\kappa}^* (MM^*)^{\dagger} [M]_{\kappa} = [Y]_{\kappa\kappa}, \tag{15}$$

where  $[Y]_{\kappa\kappa}$  is the  $\kappa$ -th diagonal element of  $Y$ . Since  $Y$  is equal to  $Id$  in the ideal case, hence  $[Y]_{\kappa\kappa}$  should be equal to 1. We consider the error-type characteristic:

$$E_{\kappa}^{(i)} = 1 - [M]_{\kappa}^* (MM^*)^{\dagger} [M]_{\kappa}. \tag{16}$$

Let us consider the application of  $Y$  to the set of sources of interest  $\{[Id]_{\kappa}, \kappa \in K\}$ . For example,  $K$  can be a priori known source locations or the whole domain. We can evaluate characteristics (13), (14), and (16) on  $K$ . Corresponding cumulative characteristics are:

$$E^{(c)} = \frac{1}{|K|} \sum_{\kappa \in K} E_{\kappa}^{(c)}, \quad c \in \{n, e, i\}. \tag{17}$$

The analysis tool that does not take knowing the exact solution is the decay rate of the sensitivity operator singular spectrum  $\{s_n\}_{n=1}^N$ , as it is done in [73]. The issue with this analysis type is how to compare different singular spectra quantitatively.

### 2.5. Inverse Modeling Scenario

Constructing a realistic inverse modeling scenario is an essential part of inverse modeling. In the context of the source identification problem, the main components of an inverse modeling scenario are:

- Geographical domain.
- Monitoring system characteristics: locations and accuracy.
- Main emission sources to construct the "exact" solution.
- Chemical transformation mechanism, initial, and boundary conditions.

- Meteorological conditions, determining CTM model coefficients.

The geographical domain (Figure 1) is the most definite component of the scenario. The monitoring system is also a definite component of the inverse modeling scenario since the positions of all the monitoring sites are known. The locations of some Roshydromet sites [48] (Figure 1) were taken as the prototype of the measurement system in the scenario. According to [48], the system measures several components, including  $SO_2$ ,  $CO$ ,  $NO_2$ ,  $NO$ ,  $O_3$ ,  $HS$ ,  $PM_{2.5}$ , and  $PM_{10}$ . The measurement data is provided in the open-access mode on a daily-averaged basis. In the scenario, we suppose that there are only  $O_3$  concentration measurements ( $L_{meas} = \{\#(O_3)\}$ ). Here  $\#(O_3)$  denotes the index of  $O_3$  concentration in the state function  $\varphi$ . The types of measurements collected at the measurement sites are defined in the numerical experiments.

In the considered scenario, the images (Snapshots) cover the whole domain simultaneously. The geostationary satellites can potentially obtain this type of image [74,75]. Nevertheless, the instant image of the whole domain is a simplification since each instrument has its spatial and temporal resolution and coverage. Non-geostationary satellites collect images as the series of smaller images along their tracks with a given swath width [76–78]. More details about the satellite images of the concentration fields can be found in [74–78]. In our scenario, the images of the concentration field are available on the third day of the period ( $t = 3 \times 24 \times 3600$  s).

Emission sources are the factor that largely determines the air quality in the region, but the information about the sources is incomplete. This is the main reason for developing the source identification algorithms. We need to construct a realistic “exact” solution to move in this direction. It will be required to check the source identification algorithms (see Section 2.4.1) and to provide the estimate of  $\mathbf{q}^{(*)}$  to evaluate the sensitivity operator  $M_U[\mathbf{q}^{(*)}, \mathbf{q}]$  for the analysis purposes (see Section 2.4.2).

We construct the “exact” solution as a set of Pointwise sources with constant time emission rates. Furthermore, we suppose that the only emitted substance is  $NO$  ( $L_{src} = \{\#(NO)\}$ ). In this scenario, we have so-called “indirect” measurements, when the measured substance ( $O_3$ ) is different from the emitted one ( $NO$ ).

To define realistic emission rates, we depart from the estimate  $\bar{q}_{NO}^{(Irkutsk)}$  of the Irkutsk (the largest city in the region) total production of  $NO$  from [65]:

$$\bar{q}_{NO}^{(Irkutsk)} = 125 \times 10^3 \frac{mg}{s}. \tag{18}$$

To obtain the concentration change in the grid cell due to the emissions, we divide  $\bar{q}_{NO}^{(Irkutsk)}$  by the volume of the grid cell  $V_{cell}$  in  $m^3$ :

$$V_{cell} = 19502 \times 21991.6 \times 100m^3, \\ q_{NO}^{(Irkutsk)} \approx 2.91 \times 10^{-6} \frac{mg}{s \ m^3}.$$

In the figures, we use the dimensionality  $\frac{mg}{s \ m^3}$  for the sources and  $\frac{mg}{m^3}$  for the concentrations.

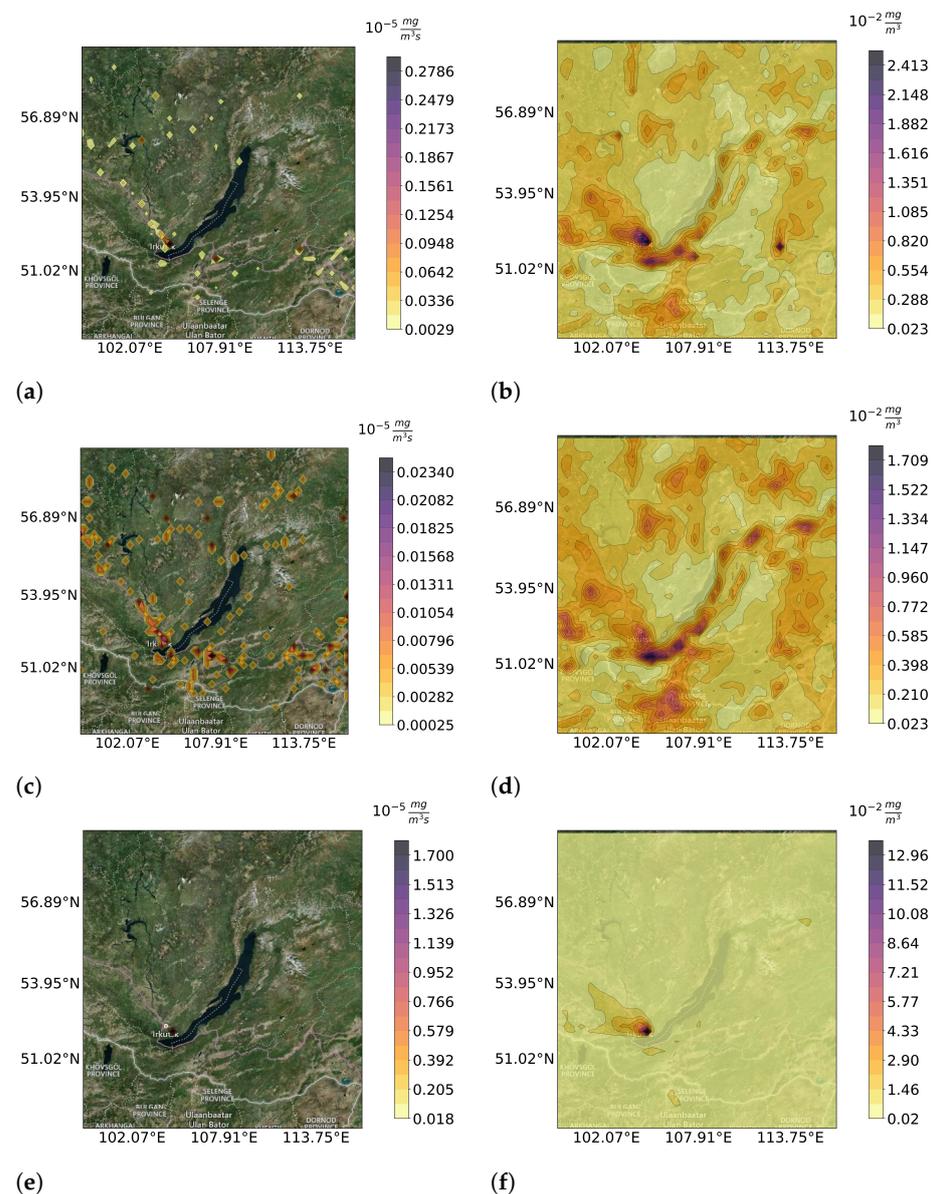
The emission of the  $c$ -th city in the region is estimated as proportional to the ratio of its population to Irkutsk’s population:

$$q_{NO}^{(c)} = q_{NO}^{(Irkutsk)} \times \frac{Population^{(c)}}{Population^{(Irkutsk)}}, \tag{19}$$

where  $Population^{(c)}$  is the population in the  $c$ -th city according to Wolfram Alpha Service [79]. This construction of the scenario allows obtaining a spatially heterogeneous distribution of the unequal emission sources.

The city’s emissions are attributed to the nearest grid-point. If more than one city corresponds to a grid point, then the emissions of these cities are added. We call this configuration a “realistic” source configuration (Figure 2a). In the second configuration

(Figure 2c) all the cities have identical emission rates, while the total of the emission rates is the same as in the “realistic” source configuration. We call it a “unified” configuration. In the third configuration (Figure 2e) all the emissions from the “realistic” configuration are localized in the nearest grid-point to the Irkutsk city. We denote this configuration by a “single”. All these source configurations have the same  $L_1$  norm. The initial guess for the source identification is zero emissions. Therefore, the relative error of the initial guess is 100%. A priori emission source is considered as zero  $f = 0$ .



**Figure 2.** Emission sources and corresponding mean concentration fields of NO for considered source configurations: “realistic” (a,b); “unified” (c,d); “single” (e,f), respectively.

Along with the anthropogenic sources, the sources and sinks are determined by the natural processes: rains and snowfalls, vegetation, forest fires, swamps, etc. The chemical reactions, unconsidered in the transformation model, also contribute to the distributed sources and sinks. These sources are not addressed in the paper since we use only synthetic data. In the case of the real measurement data, these sources should also be taken into consideration. The primary anthropogenic emission sources and the natural factors define the chemical transformation model to be considered. Since the inverse modeling tasks are

time-consuming, the chemical system should be as small as possible but relevant to the main chemical components, especially the major primary and secondary pollutants.

As the transformation model (operators  $P$  and  $\Pi$  in (1)), we chose the Leighton relationship-based atmospheric chemistry model (e.g., [80]), described in Appendix C. Initial conditions  $\varphi_l^0$  are taken as constant in space according to (A11)–(A13). This simplification can be refined by using the output of the global chemical transport models (e.g., [81]) or setting the model’s spin-up period. The boundary conditions are:

$$\alpha_l = \begin{cases} 0, & (\mathbf{x}, t) \in \Gamma^{(out)} \\ \varphi_l^0, & (\mathbf{x}, t) \in \Gamma^{(in)}, \end{cases} \quad \beta_l = 0, \tag{20}$$

where  $\varphi_l^0$  is the constant value of the corresponding initial conditions. The values of  $\varphi_l^0$  denote background concentrations of the considered substances. This simplification can also be refined by taking the output of the global CTMs as the boundary conditions.

Meteorological conditions are the most variable component in the system. Modern meteorological models can evaluate the meteorological fields with high accuracy. The weather data for the scenario was chosen for the period from 12:00 on 23.07.2019 to 12:00 on 26.07.2019. The results of the COSMO (Consortium for Small-scale Modeling) model calculation [82] are used as the 2D wind-speed vector field  $\mathbf{u}$  at the surface level. The description of the meteorological conditions is given in Appendix D. The diffusion coefficient  $\mu = 1000 \text{ m}^2/\text{s}$  was chosen as a constant one.

To generate the “exact” solution  $\varphi^{(*)}$ , we solve the direct problem. The  $\text{O}_3$  elements of the “exact” solution are used to obtain the synthetic measurement data. To show the differences between the source configurations, we present mean (over time interval  $[0, T]$ )  $\text{NO}$  concentration fields in Figure 2b,d,f.

### 3. Results

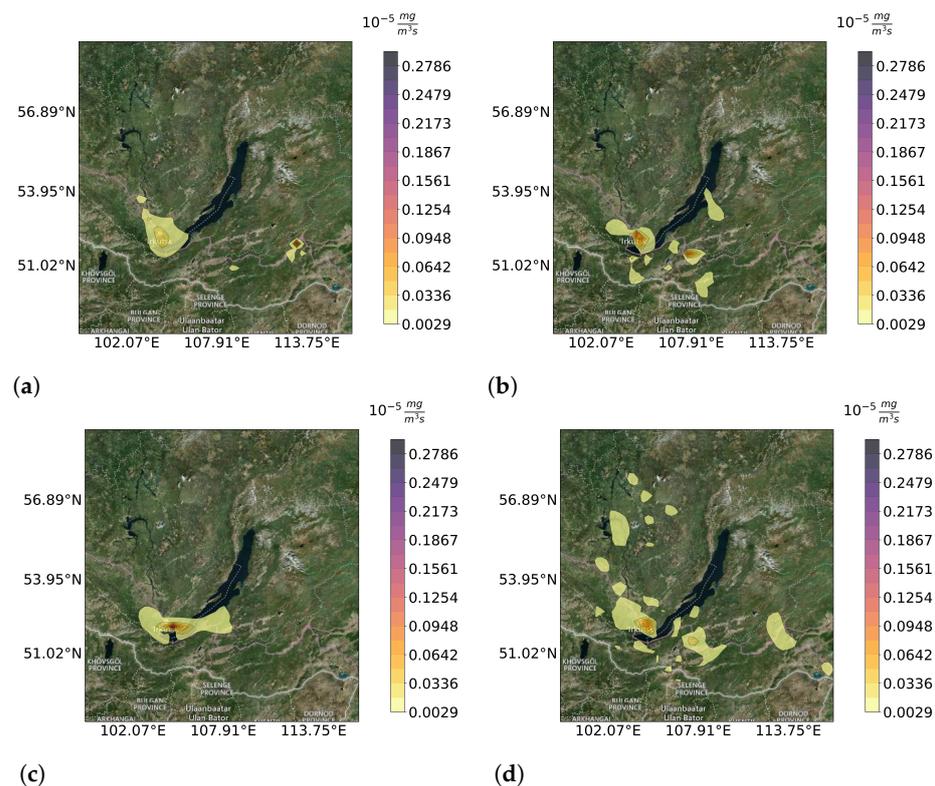
#### 3.1. Heterogeneous Measurements

To evaluate the synergistic effect of using different measurement types, we compare the source identification (or simply reconstruction) results obtained for composite measurements with those obtained for the specialized measurement types. The measurement sites are divided into three groups (Figure 1): Timeseries, Pointwise, and Integral. In the considered configurations, we use the parameters of the projection functions ensembles given in Table 1. Pointwise measurements are taken at 6-hour time intervals.

**Table 1.** Adjoint ensemble sizes.

Type	$\Xi_{Type}$	$N_{Type}$	Description
Pointwise	60	60	5 sites $\times$ 12 measurement moments
Timeseries	60	6	6 sites $\times$ 10, $\Theta_{Timeseries} = 10$
Integral	5	5	5 sites
Snapshot	625	1	1 image $\times$ 25 $\times$ 25, $\Theta_{Snapshot}^{(x)} = \Theta_{Snapshot}^{(y)} = 25$
Composite	750		Sum of the above

The source reconstruction results for individual measurement types in the “realistic” source configuration are presented in Figure 3. We see that the positions of “Pointwise” measurements are far from the largest source cluster located near Irkutsk. Respectively, the reconstruction by “Pointwise” measurements is less accurate for this cluster compared to other measurement types. Snapshot reconstruction allows one to identify the sources that are far from the air quality monitoring network. Still, it contains many artifacts explainable by a relatively small number of the projection functions.

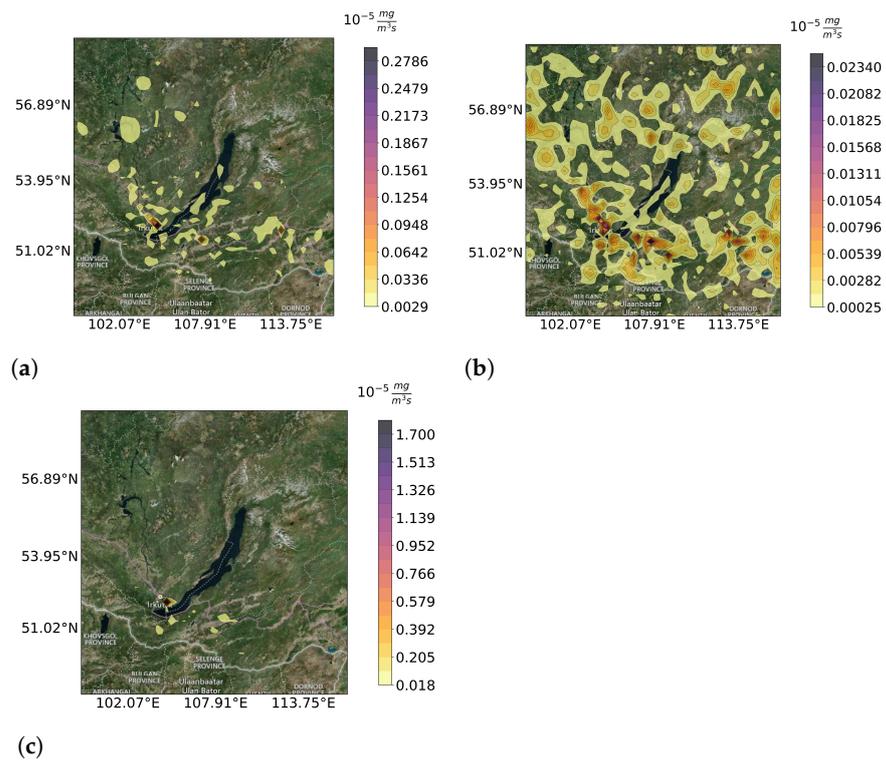


**Figure 3.** Source reconstruction results in the “realistic” source configuration for specific measurement types: Pointwise (a); Timeseries (b); Integral (c); Snapshot (d).

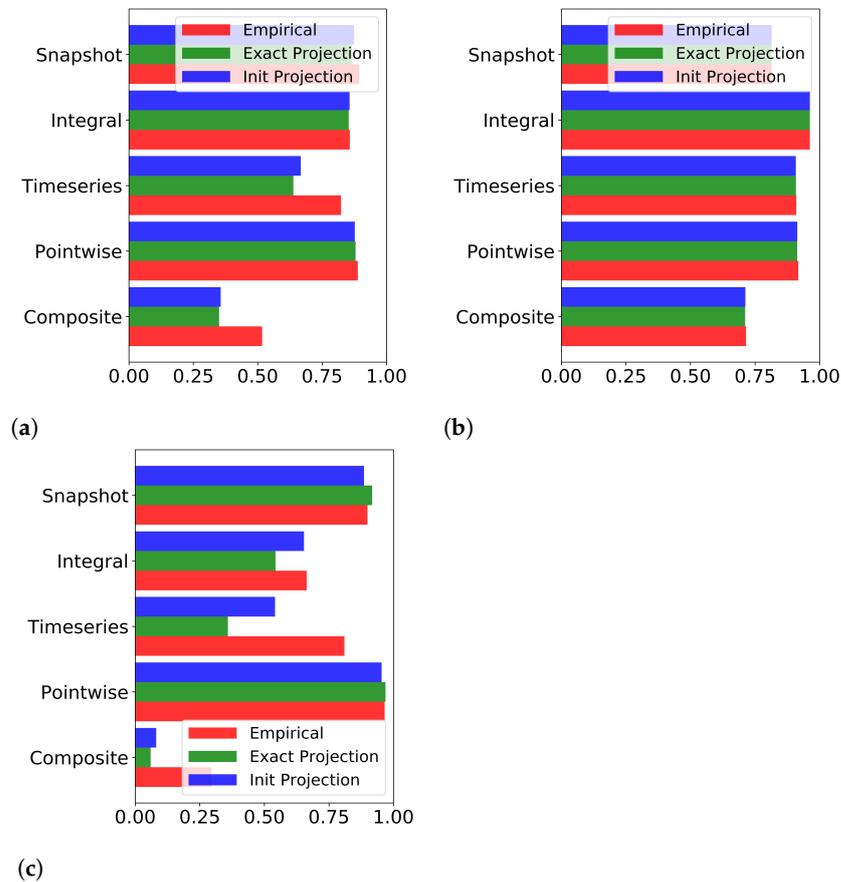
In Figure 4, we compare composite data reconstruction results for different source configurations. In Figure 5, the relative errors in different measurement configurations and their estimates are presented. Here “empirical” denotes the error obtained with the inverse problem solution algorithm; “init projection”, and “exact projection” denote the errors of the estimates (11) and (12), respectively. The less is the red (“empirical”) bar, the better the solution is. Since “init projection” (blue bars) and “exact projection” (green bars) are the estimates of the “empirical” relative error, then in the ideal case, they should have the same lengths as the red bar. As we can see, projections (11) and (12) give a relatively accurate estimate of the inverse problem solution.

Comparing the different source configuration results, we can see that the best reconstructions for the composite measurements are obtained for “realistic” and “single” emission sources. The best “Snapshot” results are obtained for the “unified” sources configuration. The localized measurements (Pointwise, Timeseries, Integral) are taken near the most populated cities in the region. In the “realistic” and “single” source configurations, they provide better reconstruction than in the “unified” case. In the “unified” case, the location of the measurements plays a less important role, and the “Pointwise” reconstruction results are comparable to the “Timeseries” results.

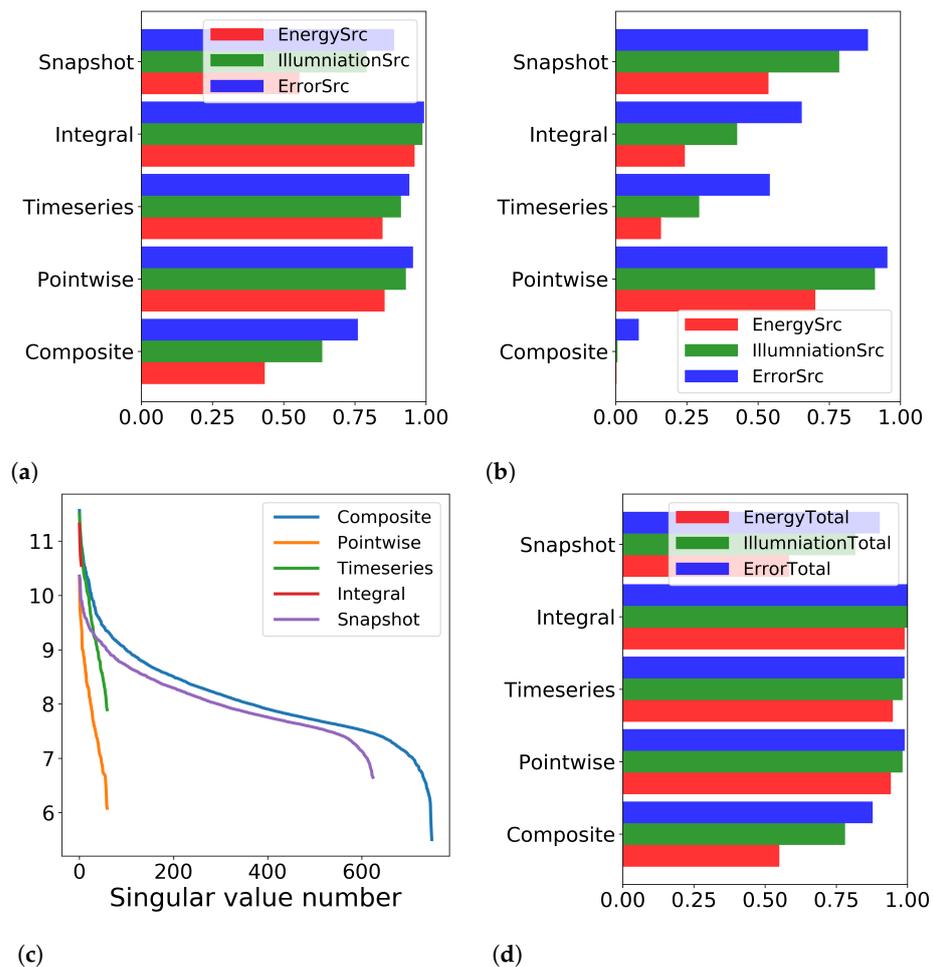
In Figure 6, we compare the characteristics that do not know the exact emission rates. Here “EnergySrc”, “ErrorSrc”, and “IlluminationSrc” denote  $E^{(e)}$ ,  $E^{(n)}$ , and  $E^{(i)}$  from (17), respectively, for  $K$  containing the indices of the grid points with “cities”. “EnergyTotal”, “ErrorTotal”, and “IlluminationTotal” denote the same characteristics but for  $K$  containing all grid points indices. These characteristics are also considered as the estimates of the “empirical” relative error. The characteristics in Figure 6a,b take the information of the source positions. The characteristics in Figure 6c,d are completely independent of the information about the sources.



**Figure 4.** Composite data reconstruction results for different source configurations: “realistic” (a); “unified” (b); “single” (c).



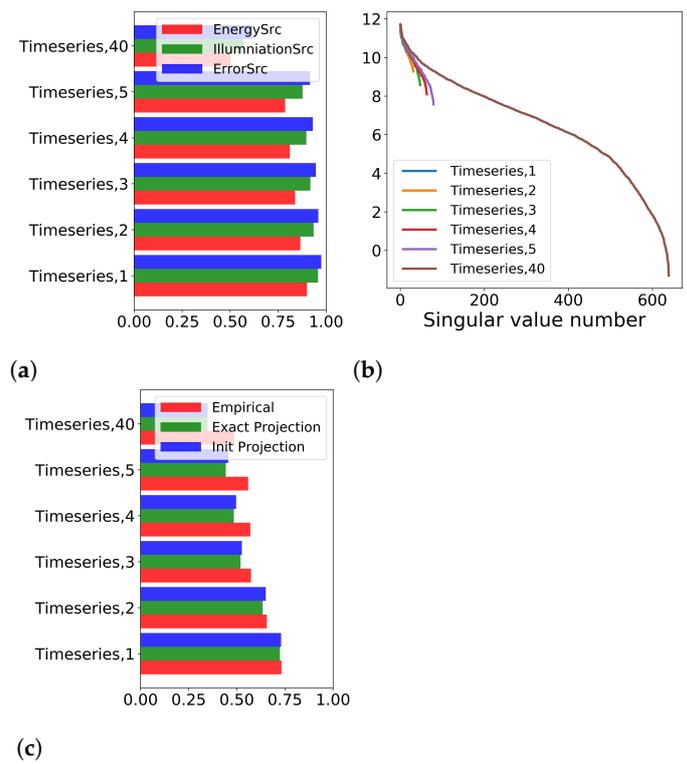
**Figure 5.** Comparison of empirical reconstruction errors and their projection-based estimates for “realistic” (a); “unified” (b); “single” (c) sources. These estimates take knowing the “exact” solution.



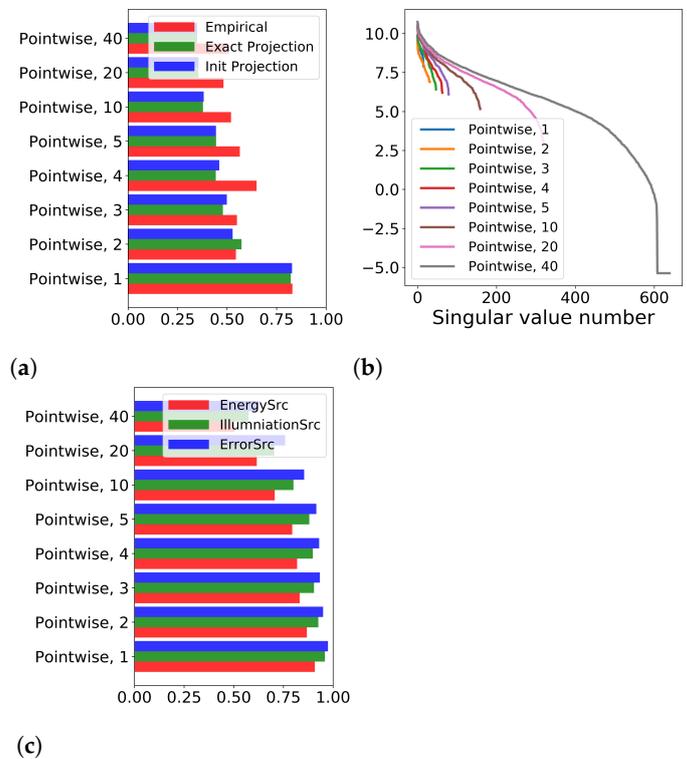
**Figure 6.** Characteristics dependent on the source locations for “realistic”, “unified” (a), and “single” (b) source configuration. Characteristics that are independent of the source locations:  $\text{Log}_{10}$  of the sensitivity operator  $M_U[\mathbf{q}^{(0)}, \mathbf{q}^{(0)}]$ 's matrix singular spectrum versus the singular value number (c), and the same characteristics as in Figure 6a,b, calculated for the whole domain (d).

### 3.2. Specific Measurement Types

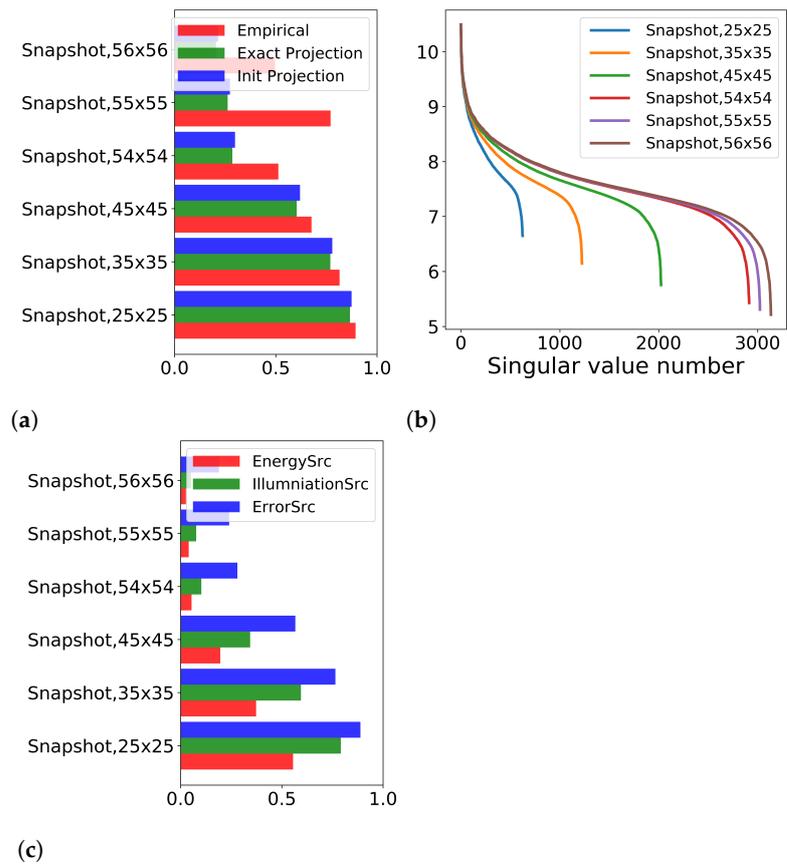
In these experiments, we compare the reconstruction results for the “realistic” source configuration in cases when the whole measurement network collects data of the same type. The second question is how the ensemble parameters (such as  $\Theta_{\text{Timeseries}}$ ,  $\Theta_{\text{Snapshot}}^{(x)}$ ,  $\Theta_{\text{Snapshot}}^{(y)}$ ) influence the reconstruction results. In Figure 7, we present the results for different numbers of projection functions in the case of time-series measurements. The label on a figure “Timeseries,  $N$ ” means that  $\Theta_{\text{Timeseries}} = N$ . In Figure 8, we compare the results of Pointwise measurements for different numbers of measurement acquisition times. The label on a figure “Pointwise,  $N$ ” means that time series in each of 16 sites are approximated by  $N$  values uniformly distributed on the time interval, i.e., for “Pointwise, 40”  $\Xi = 16 \times 40$ . In the last experiment (Figure 9), we consider reconstruction results for the different number of projection functions in the Snapshot case. The label on a figure “Snapshot,  $N \times N$ ” means that  $\Theta_{\text{Snapshot}}^{(x)} = N$ ,  $\Theta_{\text{Snapshot}}^{(y)} = N$ . In Figure 10, we compare the reconstruction results for different numbers of the projection functions in the Snapshot scenario.



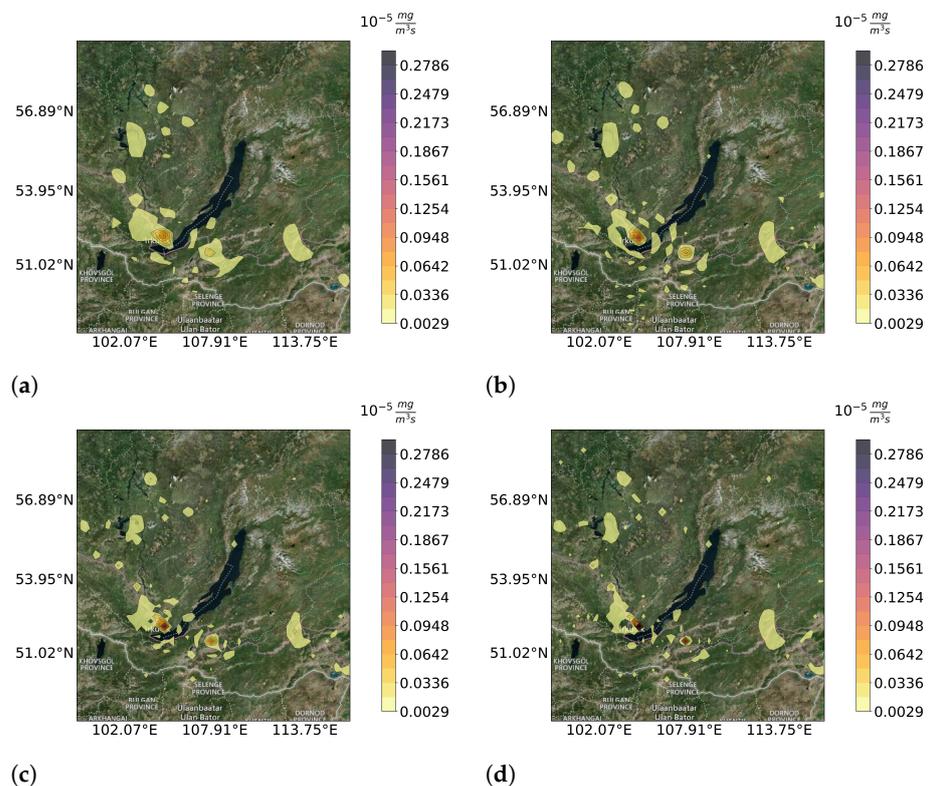
**Figure 7.** Reconstruction with Timeseries measurements for different numbers of projection functions. “Empirical” reconstruction error and error estimated by projection (a);  $\text{Log}_{10}$  of the sensitivity operator’s singular values versus the singular value number (b); Projection operator characteristics (c).



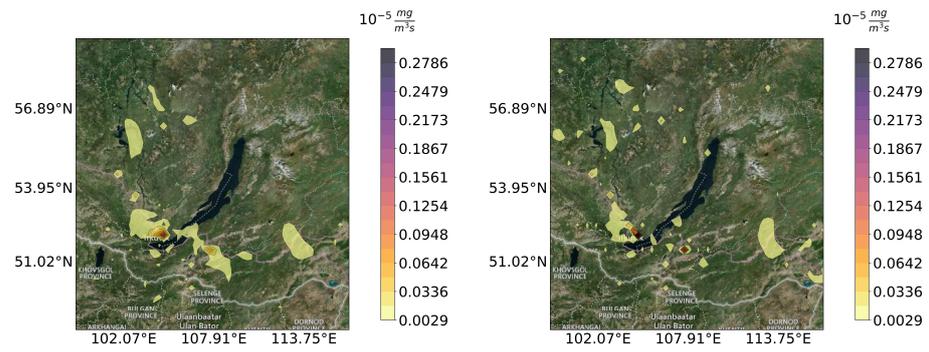
**Figure 8.** Reconstruction with Pointwise measurements for different numbers of measurement times. “Empirical” reconstruction error and error estimated by projection (a);  $\text{Log}_{10}$  of the sensitivity operator’s singular values versus the singular value number (b); Projection operator characteristics (c).



**Figure 9.** Reconstruction with Snapshot measurements for different numbers of projection functions. “Empirical” reconstruction error and error estimated by projection (a);  $\text{Log}_{10}$  of the singular spectrum versus the singular value number (b); Projection operator characteristics (c).

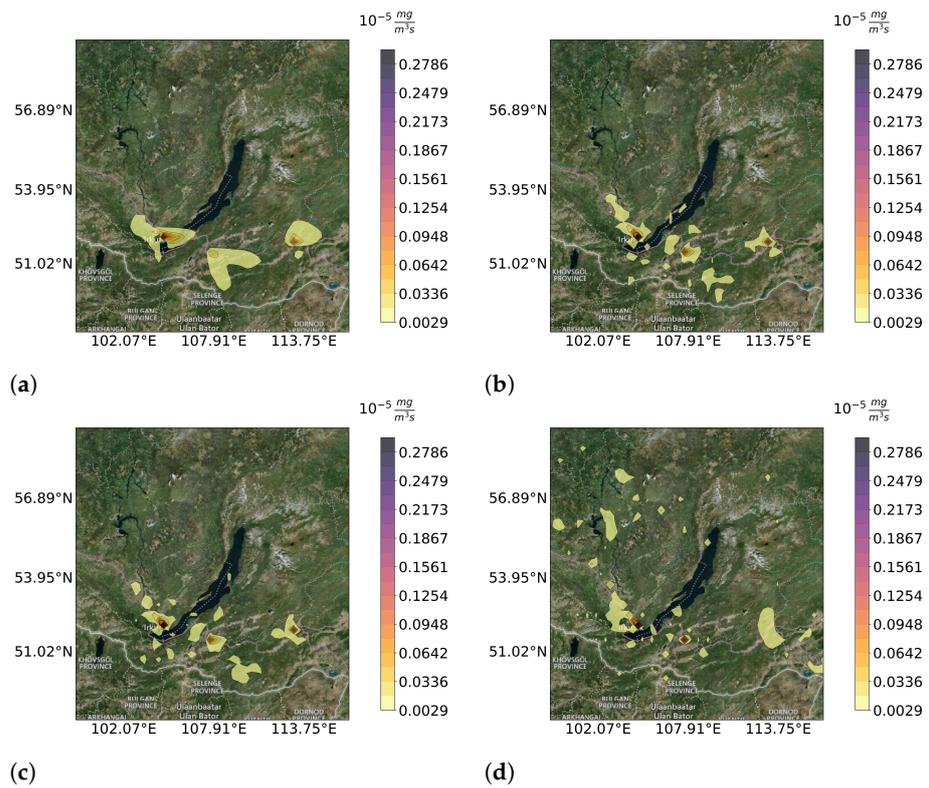


**Figure 10.** Cont.

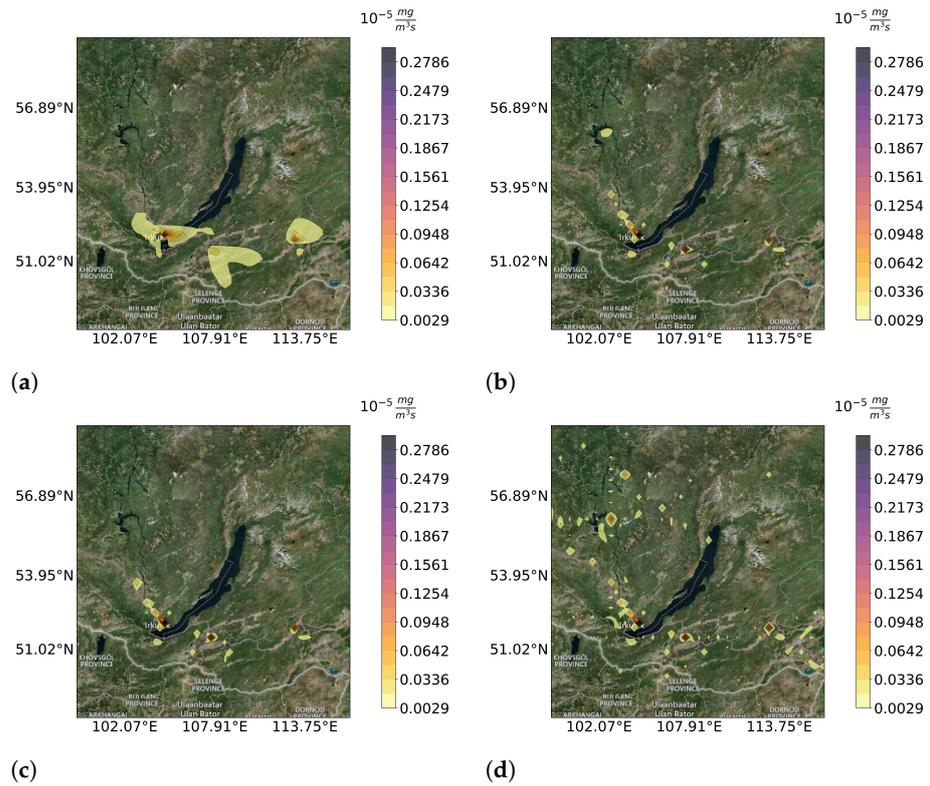


(e) (f)  
**Figure 10.** Snapshot reconstruction for different numbers of ensemble members:  $25 \times 25$  (a);  $35 \times 35$  (b);  $45 \times 45$  (c);  $54 \times 54$  (d);  $55 \times 55$  (e);  $56 \times 56$  (f).

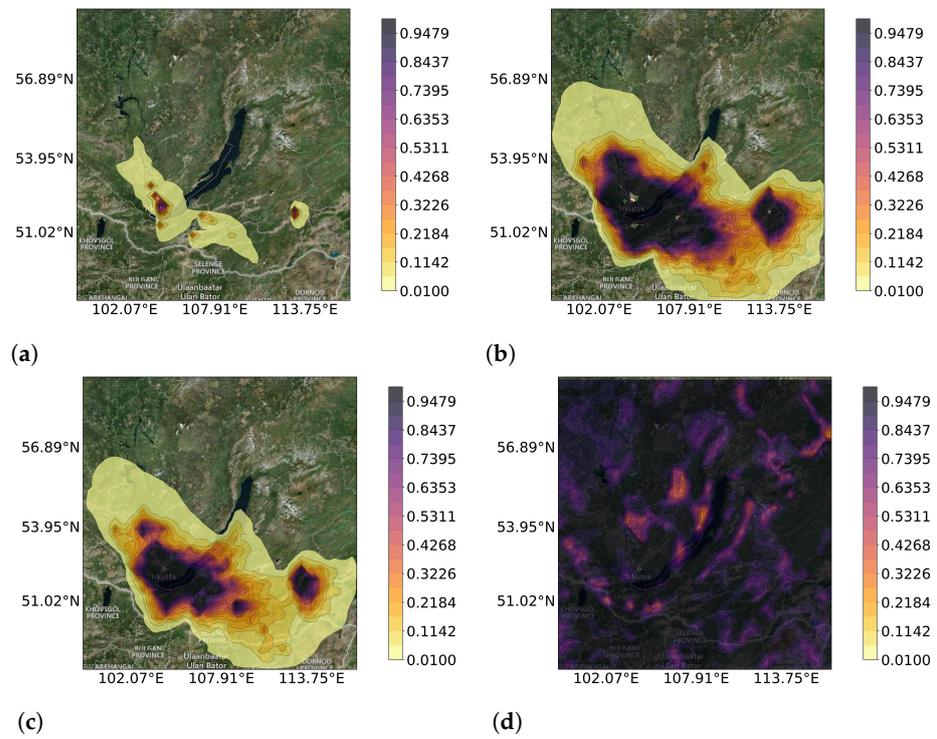
In Figure 11, we compare the reconstruction results for different measurement types. Timeseries configuration is presented for  $\Xi = 40 \times 16$ , Pointwise for  $\Xi = 40 \times 16$ , and Snapshot for  $\Xi = 56 \times 56$ . In Figure 12, we can see the projection-based estimates of the results in Figure 11. In Figure 13, there are the corresponding illumination functions (15).



(a) (b)  
 (c) (d)  
**Figure 11.** Reconstruction results for different measurement types: Integral (a); Timeseries (b); Pointwise (c); Snapshot (d).



**Figure 12.** Initial guess-based projection estimate (11) for different measurement types: Integral (a); Timeseries (b); Pointwise (c); Snapshot (d).



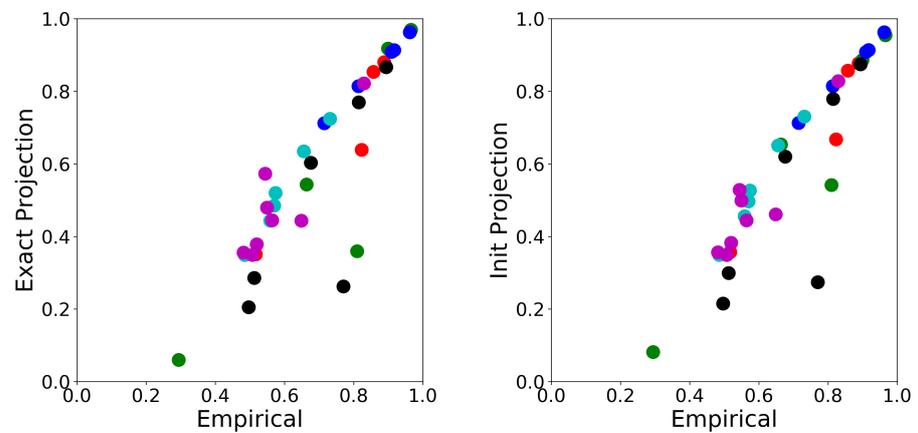
**Figure 13.** Illumination functions (15) for different measurement types: Integral (a); Timeseries (b); Pointwise (c); Snapshot (d).

### 3.3. Accuracy of the Reconstruction's Prediction

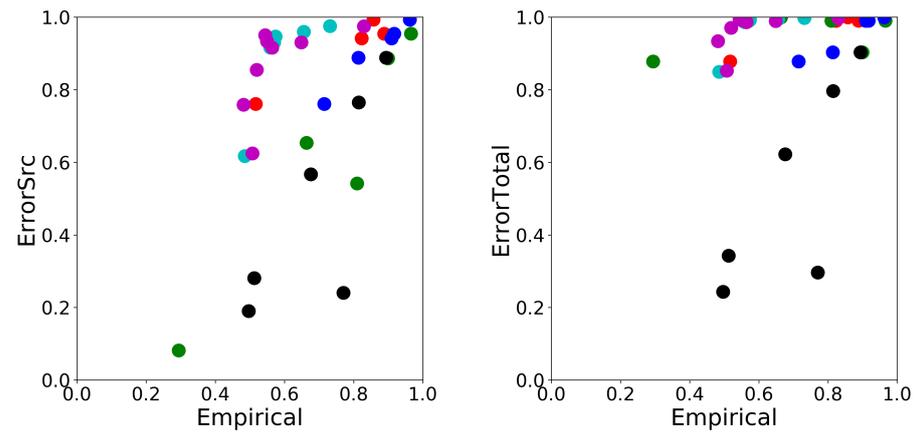
To summarize the relationship between the considered characteristics and the resulting “Empirical” source identification errors, we made the scatter plots in Figures 14 and 15. The colors in Figures denote different experiments:

- Red: Section 3.1, “realistic” source case;
- Green: Section 3.1, “single” source case;
- Blue: Section 3.1, “unified” source case;
- Black: Section 3.2, Timeseries experiment;
- Cyan: Section 3.2, Pointwise experiment;
- Magenta: Section 3.2, Snapshot experiment.

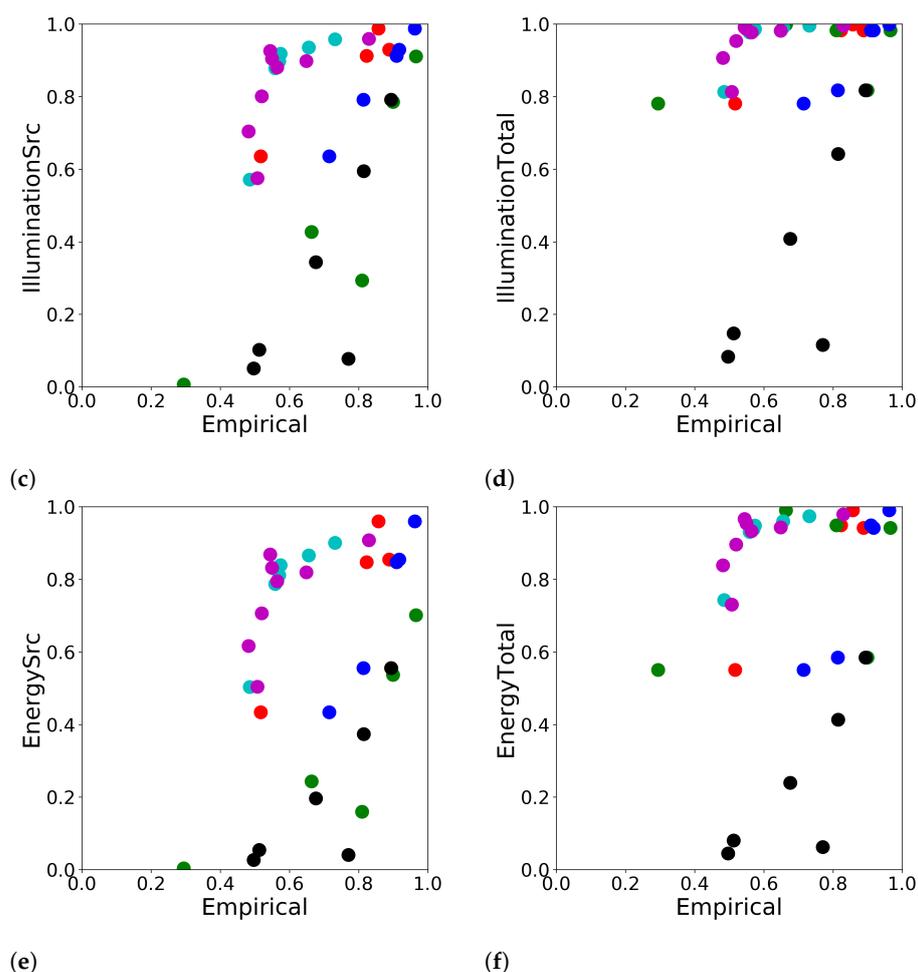
In Figure 14, we present the characteristics that take the “Exact” solution. The characteristics in Figure 15a,c,e take the locations of the “Exact” sources. The characteristics in Figure 15b,d,f are entirely independent of the “Exact” solution.



(a) (b)  
**Figure 14.** Correspondence of the “Empirical error” to “Exact Projection” (a) and “Init Projection” (b) in the numerical experiments. The different colors denote different experiments.



(a) (b)  
**Figure 15.** Cont.



**Figure 15.** Correspondence of the “Empirical error” to “ErrorSrc” (a), “ErrorTotal” (b), “IlluminationSrc” (c), “IlluminationTotal” (d), “EnergyTotal” (e), and “EnergySrc” (f) in the numerical experiments. The different colors denote different experiments.

#### 4. Discussion

An essential advantage of the sensitivity-operator-based approach to inverse modeling is that various problems can be reduced in a unified way to a family of quasi-linear operator equations. Due to the properties of adjoint equations and sensitivity relations, the developed approach makes it possible to work in a unified way with heterogeneous measurement data. Note that these properties of the sensitivity relations are also used in the variational algorithms where various measurements can be combined by adding corresponding discrepancy functions to the cost function. However, in contrast, the sensitivity-operator-based approach also provides a straightforward way to analyze the inverse problem properties by analyzing the properties of the corresponding sensitivity operator family. Moreover, a sensitivity operator can combine not only different measurement types and their ensembles. If the source (or source pattern) persists for a long time, it may combine different meteorological scenarios to a single sensitivity operator.

Sensitivity functions are calculated with the adjoint equation’s solutions. The adjoint equation structure is determined by the equations in variations for the direct problem, i.e., by the process model. We do not discuss the limitations of the approach that are inherited from the process model here (1)–(4), since the air quality model development is a separate work direction, which deserves special attention. Nevertheless, the need to derive the adjoint problems for complicated models can also be considered as the limitation of the approach. This limitation can be alleviated by using automatic adjoint code generators [83–85].

In turn, the measurement data determine the sources for the adjoint equations. Hence, a limitation of the approach is that a measurement data element has to be presented as the application of the scalar-product-type aggregate to the model state function. This is a weak limitation in the case of the linear measurement operators, connecting the model's state function with the measurement data. To consider nonlinear measurement operators, the adjoint problem has to be altered. This is a matter of further research.

Based on the sensitivity operator analysis of the inverse problem, methods of express analysis and refined analysis of the information content of monitoring systems data can be considered. The quasi-linear structure of the resulting operator equation admits using linear (matrix) analysis tools to estimate the inverse problem properties. Since solving the inverse problem may take a significant amount of time compared to the solution of the direct problem, it is advantageous to have an express estimate of the inverse problem solution quality that can be achieved in the given conditions. This is especially important for the monitoring system design (or optimization) task when one has to try and compare different system configurations. Our approach gives such a possibility straightforwardly.

On the other hand, the gain in computational time takes its payback. The difference between the "empirical" relative error and its estimates can be considered as the limitation of the approach. To overcome this limitation, we recommend checking the conclusions obtained with the estimates by the "empirical" solution of the inverse problem. There can be several reasons for these differences. The first one is that the properties discussed in Section 2.4.2 are obtained for the linear approximates of the sensitivity operator. The quasi-linear construction of the sensitivity operator admits the use of higher-order analysis and solution algorithms. It may be the direction of future work. Nevertheless, we can see that (11) and (12) provide similar estimates (Figures 5, 7a, 8a, and 9a) notwithstanding that (12) uses the exact solution in the sensitivity operator and (11) does not. It may indicate that some inaccuracy in the sensitivity operator can be tolerated when using it to make projection-type estimates.

The second possible reason for the differences is that the considered sensitivity operator properties are independent of the operator equation (8) solution algorithm (e.g., its parameters and initial guess). At the same time, the "empirical" results are determined by the algorithm's efficiency.

In the ideal case, the points in Figures 14 and 15 should lie on the diagonal. This configuration would mean that the estimates match the "empirically" ("experimentally") obtained errors. Another favorable situation is when the estimate depends on the error monotonically. This dependence allows using such estimates in the observation system optimization. Analyzing Figures 14 and 15, we conclude that projection-based estimates, summarized in Figure 14, provided the most efficient estimates of the source identification problem. The estimates in Figure 15a,c,e are harder to interpret in the source identification error terms. We can tentatively conclude that there may be some monotonic relation between the predicted and predicting characteristics in the given experiments. Moreover, we can conclude that there is a difference between the "ErrorSrc", "IlluminationSrc", and "EnergySrc" characteristics, but it is not essential. The relation between the characteristics in Figure 15b,d,f is the most unclear, and currently, we can not recommend using it as the estimate.

We did not include a priori information about the Pointwise character of the sources in the algorithm. On the one hand, the account of this information can reduce the problem's uncertainty, but on the other hand, there can be the sources treated as distributed ones. For example, the sources, such as intercity highways, may be considered linear sources. The chemical reactions and aerosol formation processes not considered or improperly considered in the model can be responsible for spatially distributed emission sources or sinks. Moreover, rains, snowfalls, forests, swamps, lakes, and rivers can also be responsible for spatially distributed sources and sinks. See Appendix D for the description of the weather conditions in the specified time interval.

In this study, we mainly considered the spatial and temporal resolution of the measurement data. Another essential question, especially important for heterogeneous measurement data, is different measurement data quality. To take into account these differences, we can weigh corresponding projection functions accordingly. The analysis of the different data quality in the case of heterogeneous measurements is a matter of future research.

As we see, the reconstruction results with respect to “exact” solutions have distinct artifacts, especially in the case of small projection function sets. In the current paper, we deal with this issue by increasing the adjoint ensembles, which comes at the price of increasing the computation time. The need for large ensembles can be considered as the limitation of the approach. This limitation is weakened by the fact that the sensitivity operator construction takes the solution of ensembles of similar but independent adjoint problems. Therefore, the elaborated algorithm may be naturally parallelized. Previously we carried out such an analysis in [86].

To further improve the technology, we plan to test machine learning approaches [87–89] to reduce these identification artifacts for relatively small ensembles. Another way to optimize (reduce) the number of the ensemble elements is to evaluate the most informative ensemble members (as it was done in [42] by choosing the elements with the largest projection on the initial discrepancy) or aggregate the ensemble members (as it was done in [90], where the elements were aggregated according to the left singular vectors of the sensitivity operator). This is a matter of separate research, which is justified by the current results demonstrating the potential advantages of the approach.

## 5. Conclusions

The combined use of process models and observational data makes it possible to solve a number of environmental problems, including the identification of pollution sources. Moreover, the mathematical model and source identification problem can integrate heterogeneous air quality measurement data.

The sensitivity-operator-based approach provides tools with different information requirements and computation complexity and intensity for analyzing air quality monitoring systems. The approach was illustrated by analyzing the monitoring system in the Lake Baikal region. As the result of the analysis, we estimated which of the regional source clusters are observable by the monitoring systems with different configurations and data acquisition schemes.

In conclusion, we note the approach can be applied to the other sets of process models and believe the general inverse modeling scheme based on the sensitivity operators will successfully solve a wide range of environmental protection, atmosphere, and ecology problems.

**Author Contributions:** Conceptualization, A.P. and V.P.; methodology, A.P. and V.P.; software, A.P. and A.G.; validation, A.P., E.P., and V.K.; formal analysis, A.P.; investigation, A.P. and E.T.; resources, A.G.; data curation, A.G.; writing—original draft preparation, A.P.; writing—review and editing, A.P., E.T., E.P., and A.G.; visualization, A.P. and A.G.; supervision, V.P.; project administration, A.P.; funding acquisition, A.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was supported by the grant 075-15-2020-787 in the form of a subsidy for a major scientific project from the Ministry of Science and Higher Education of Russia (project “Fundamentals, methods and technologies for digital monitoring and forecasting of the environmental situation on the Baikal Natural Territory”).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available on request due to the large amount and computational nature.

**Acknowledgments:** We thank Igor Chernikh for providing technical support, as well as for consulting in choosing efficient compiler optimization options. We are grateful to Valery Tokarev for the synoptic

analysis of the experiment dates. The initial data of the global ICON model (DWD) were used (<https://opendata.dwd.de/weather/nwp/icon/grib>, accessed on 23 July 2020). Observations from the synoptic station were obtained through the open-access NOAA FTP server.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**Abbreviations**

The following abbreviations are used in this manuscript:

- BNT        Baikal Natural Territory
- SVD        Singular value decomposition
- COSMO    Consortium for Small-scale Modeling
- CTM        Chemical transport model

**Appendix A. Newton–Kantorovich-Type Algorithm**

To solve (8) in the numerical experiments, we use the Newton–Kantorovich-type Algorithm A1, which is the modified version of the algorithm from [41]. Equation (8) is presented in the form:

$$M_U[\mathbf{q}, \mathbf{q}] (\mathbf{q}^{(*)} - \mathbf{q}) = H_U I - H_U \boldsymbol{\varphi}[\mathbf{q}] + (M_U[\mathbf{q}, \mathbf{q}] - M_U[\mathbf{q}^{(*)}, \mathbf{q}]) (\mathbf{q}^{(*)} - \mathbf{q}) + H_U \delta I.$$

The inversion procedure that is used in Newton-type algorithms is regularized with truncated SVD:

$$\Theta(\mathbf{m}, \Sigma) := \mathbf{m}^T [\mathbf{m}\mathbf{m}^T]_{\Sigma}^+ . \tag{A1}$$

Here  $[\mathbf{m}\mathbf{m}^T]_{\Sigma}^+$  denotes the regularized matrix inversion procedure based on the truncated SVD. We need the following notation for the r-pseudoinverse matrix [91] for a matrix  $\mathbf{m}\mathbf{m}^T \in \mathbb{R}^{K \times K}$ :

$$[\mathbf{m}\mathbf{m}^T]_{\Sigma}^+ = \sum_{l=1}^p \frac{\mathbf{U}_l}{s_l^2} \langle \cdot, \mathbf{U}_l \rangle_{\mathbb{R}^K}, \quad s_1^2/s_p^2 \leq \Sigma < s_1^2/s_{p+1}^2, \tag{A2}$$

where  $\langle \cdot, \cdot \rangle_{\mathbb{R}^K}$  is the Euclidean scalar product in  $\mathbb{R}^K$ ,  $\{\mathbf{U}_l\}_{l=1}^{rank(\mathbf{m})}$  is the orthonormal system of left singular vectors of  $\mathbf{m}$ , and  $s_l$  are the singular values.

The first modification of Algorithm A1 with respect to the algorithm from [41] is that the function  $J_U$  is enforced to decrease monotonically by the choice of the appropriate step parameter  $\gamma$  as in [92]:

$$J_U(\mathbf{q}) := \|H_U I - H_U \boldsymbol{\varphi}[\mathbf{q}]\|_{\mathbb{R}^E}^2. \tag{A3}$$

The second modification is the absence of the projection to the set  $Q$  after a Newton-type step.

Algorithm A1 has three levels of nested iterations. On the upper level, the inversion regularization parameter  $\Sigma$  is increased. On the second level, the Newton-type iterations are carried out up to the stabilization with the fixed  $\Sigma$ . On the third level, the step parameter  $\gamma$  is chosen to provide the monotonic decrease of the data misfit. In order to initialize the algorithm, one has to set  $\Delta\Sigma$ ,  $\mathbf{q}_0$ ,  $\varepsilon_{stab}$ , and  $i_{max}$ . The details can be found in [41,92].

**Algorithm A1** Newton–Kantorovich-type Algorithm

---

```

Σ ← 1
k ← 0
q(k) ← q0
Jmin ← JU(q(k))
qmin ← q(k)
while Σ < Σmax and JU(q(k)) > ||HUδI||ℝΞ2 do
  i ← 0
  repeat
    m(k) ← matrix of the sensitivity operator mU[q(k), q(k)]
    d(k) ← HUI − HUφ[q(k)]
    δq(k) ← PrsrcΘ(m(k), Σ)d(k)
    γ ← 1
    q(test)(γ) ← q(k) + γδq(k)
    while (not ||HUδI||ℝΞ2 < JU(q(test)(γ)) < JU(q(k))) do
      γ ← γ/2
    end while
    q(k+1) ← q(k) + γδq(k)
    k ← k + 1
    if Jmin ≥ JU(q(k)) then                                     ▷ Storing a “minimizing” iteration
      Jmin ← JU(q(k))
      qmin ← q(k)
    end if
    i ← i + 1
  until JU(q(k)) > ||HUδI||ℝΞ2 and ||q(k) − q(k−1)|| ≤ εstab||q(k−1)|| and i < imax
  Σ ← Σ × ΔΣ
  if Jmin < JU(q(k)) then                                       ▷ Restoring the last “minimizing” iteration
    q(k) ← qmin
  end if
end while
return qmin

```

---

**Appendix B. Projection Equivalence**

Indeed,

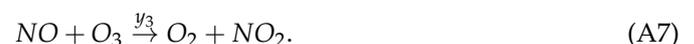
$$MM^*z = \sum_{n=1}^N U_n s_n^2 \langle U_n, z \rangle, \quad (MM^*)^\dagger z = \sum_{n=1}^N U_n \frac{1}{s_n^2} \langle U_n, z \rangle.$$

Hence,

$$M^*(MM^*)^\dagger Mz = \sum_{n=1}^N V_n \langle V_n, z \rangle. \quad (\text{A4})$$

**Appendix C. Chemical Transformation Model (Leighton Relationship-Based)**

Let us apply the developed framework to the inverse modeling scenario with a low-dimensional atmospheric chemistry model [80]:



The first reaction rate coefficient  $y_1$  is time-dependent and  $y_2, y_3$  are constant:

$$y_1(t) = \begin{cases} 10^{-5} e^{\sec(t)}, & 4 < \bar{t}_h < 20 \\ 10^{-40}, & \text{otherwise} \end{cases}, \quad (\text{A8})$$

$$\sec(t) = \left( \sin\left(\frac{\pi}{16}(\bar{t}_h - 4)\right) \right)^{0.2}, \quad \bar{t}_h = t_h - 24 \operatorname{int}\left(\frac{t_h}{24}\right), \quad t_h = \frac{t}{3600}, \quad (\text{A9})$$

$$y_2 = 1.87 \cdot 10^{-14}, \quad y_3 = 10^{-16}, \quad (\text{A10})$$

and  $\operatorname{int}\left(\frac{t_h}{24}\right)$  is the integer part of  $\frac{t_h}{24}$ .

The system's feature is that its dynamics changes dramatically when "the sun rises" at 4 o'clock. The concentration of  $O_2$  is considered constant. The initial distribution of concentrations is constant in the domain:

$$[NO] = 4 \cdot 10^{10} \frac{1}{cm^3} \approx 0.002 \frac{mg}{m^3}, \quad [NO_2] = 1.3 \cdot 10^{10} \frac{1}{cm^3} \approx 0.001 \frac{mg}{m^3}, \quad (\text{A11})$$

$$[O_2] = 5.3 \cdot 10^{18} \frac{1}{cm^3} \approx 284202 \frac{mg}{m^3}, \quad [O] = 0 \frac{1}{cm^3} \approx 0 \frac{mg}{m^3}, \quad (\text{A12})$$

$$[O_3] = 1.5 \cdot 10^{12} \frac{1}{cm^3} \approx 0.12 \frac{mg}{m^3}. \quad (\text{A13})$$

The concentrations in the model are measured in  $\frac{1}{cm^3}$ . Hence,  $q_{NO}^{(Irkutsk)}$  is equivalent to  $58493180 \frac{1}{s \cdot cm^3}$ .

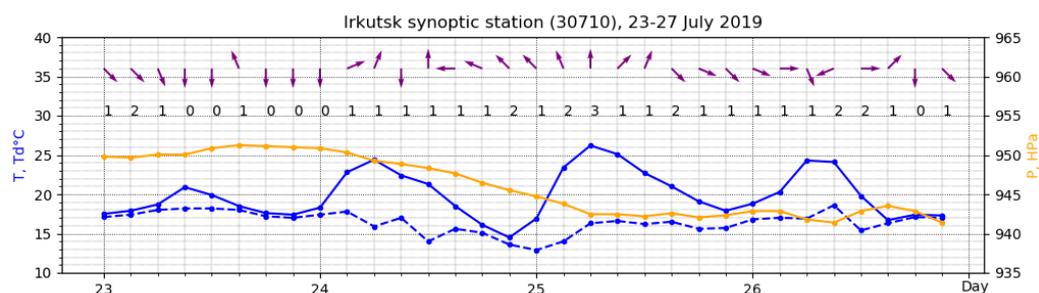
#### Appendix D. The Description of the Meteorological Scenario

According to the synoptic processes over southern Siberia–Mongolia–Transbaikalia, the period from July 23 to August 3 is generally characterized by the slow development of cyclone waves in the latitudinal band between the northern (polar) and southern (subtropical) branches of the high-altitude frontal zone (approximately between 45 and 55 degrees north latitude).

The formation of a cut-off high-altitude cyclone from the polar depression over the southern regions of Western and Eastern Siberia is best traced. In the trajectory of its blurred center, a gap is noticeable when rounding the northern spurs of the Sayan Mountains and already regenerating southeast in the Selenga basin. Heavy rainstorms in the foothills of the Altai and Sayan mountains, including the infamous Tulun flooding, are associated with this rounding and delay at the windward slopes.

Processes chronicled by date (Figure A1 shows Irkutsk observation station point):

- 23.07 Rain zone in the foothills of the Altai, in the Kuzbass. The cold front from the west offset to the east. There is practically no leading stream. Weak variable wind in the west of Lake Baikal.
- 24.07 Rain zone in the foothills of Altai-Sayan (Khakassia), Western Sayan (Daily precipitation HMS 29698 Nizhneudinsk-57mm). With the approach of a cold front from the west, the wind is mainly south-easterly.
- 25.07 The rain zone encircles the Western Sayans from the north. Cold front, offset to the east, the wind weakens and changes direction to mainly western.
- 26.07 The cold front approaches the Hangar from the west. Baikal, in the warm sector orographically cut off in the south (baric depression, thunderstorms south and north of Lake Baikal).



**Figure A1.** Observation data of wind at the 10-meter level, the temperature at the 2-meter level, and pressure at station level.

## References

- Brunet, G. *Seamless Prediction of the Earth System: From Minutes to Months*; World Meteorological Organization: Geneva, Switzerland, 2015.
- WMO. Measurement of Meteorological Variables, chapter Measurement of atmospheric composition. In *Guide to Instruments and Methods of Observation*; World Meteorological Organization: Geneva, Switzerland, 2018; Volume I, pp. 506–541.
- Penenko, V.; Raputa, V.; Panarin, A. Planning an experiment for determining the position and strength of a pollution source. *Sov. Meteorol. Hydrol.* **1985**, *11*, 10–16.
- Penenko, V.; Raputa, V.; Bykov, A. Design of an experiment for the Pollution Source Power Estimation problem. *Izv. Atmos. Ocean. Phys.* **1986**, *21*, 705–710.
- Abida, R.; Bocquet, M.; Vercauteren, N.; Isnard, O. Design of a monitoring network over France in case of a radiological accidental release. *Atmos. Environ.* **2008**, *42*, 5205–5219. [[CrossRef](#)]
- Saunier, O.; Bocquet, M.; Mathieu, A.; Isnard, O. Model reduction via principal component truncation for the optimal design of atmospheric monitoring networks. *Atmos. Environ.* **2009**, *43*, 4940–4950. [[CrossRef](#)]
- Keats, A.; Yee, E.; Lien, F.S. Information-driven receptor placement for contaminant source determination. *Environ. Model. Softw.* **2010**, *25*, 1000–1013. [[CrossRef](#)]
- Araki, S.; Iwahashi, K.; Shimadera, H.; Yamamoto, K.; Kondo, A. Optimization of air monitoring networks using chemical transport model and search algorithm. *Atmos. Environ.* **2015**, *122*, 22–30. [[CrossRef](#)]
- Ngae, P.; Kouichi, H.; Kumar, P.; Feiz, A.A.; Chpoun, A. Optimization of an urban monitoring network for emergency response applications: An approach for characterizing the source of hazardous releases. *Q. J. R. Meteorol. Soc.* **2019**, *145*, 967–981. [[CrossRef](#)]
- Kouichi, H.; Ngae, P.; Kumar, P.; Feiz, A.A.; Bekka, N. An optimization for reducing the size of an existing urban-like monitoring network for retrieving an unknown point source emission. *Geosci. Model Dev.* **2019**, *12*, 3687–3705. [[CrossRef](#)]
- Cao, S.J.; Ding, J.; Ren, C. Sensor deployment strategy using cluster analysis of Fuzzy C-Means Algorithm: Towards online control of indoor environment's safety and health. *Sustain. Cities Soc.* **2020**, *59*, 102190. [[CrossRef](#)]
- Fattoruso, G.; Nocerino, M.; Toscano, D.; Pariota, L.; Sorrentino, G.; Manna, V.; Vito, S.D.; Carteni, A.; Fabbricino, M.; Francia, G.D. Site Suitability Analysis for Low Cost Sensor Networks for Urban Spatially Dense Air Pollution Monitoring. *Atmosphere* **2020**, *11*, 1215. [[CrossRef](#)]
- deSouza, P.; Anjomshoaa, A.; Duarte, F.; Kahn, R.; Kumar, P.; Ratti, C. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustain. Cities Soc.* **2020**, *60*, 102239. [[CrossRef](#)]
- Schafer, K.; Lande, K.; Grimm, H.; Jenniskens, G.; Gijbers, R.; Ziegler, V.; Hank, M.; Budde, M. High-Resolution Assessment of Air Quality in Urban Areas—A Business Model Perspective. *Atmosphere* **2021**, *12*, 595. [[CrossRef](#)]
- Hadi-Vencheh, A.; Tan, Y.; Wanke, P.; Loghmanian, S.M. Air Pollution Assessment in China: A Novel Group Multiple Criteria Decision Making Model under Uncertain Information. *Sustainability* **2021**, *13*, 1686. [[CrossRef](#)]
- Marchuk, G. *Mathematical Models in Environmental Problems*. In *Studies in Mathematics and Its Applications Book Series*; Elsevier Science & Techn.: Amsterdam, The Netherlands, 1986; Volume 16.
- Pudykiewicz, J.A. Application of adjoint tracer transport equations for evaluating source parameters. *Atmos. Environ.* **1998**, *32*, 3039–3050. [[CrossRef](#)]
- Desyatkov, B.M.; Sarmanov, S.P.; Borodulin, A.I.; Kotlyarova, S.S.; Selegei, V.V. Determination of some characteristics of an aerosol pollution source by solving the inverse problem of pollutant spread in the atmosphere. *Atmos. Ocean. Opt.* **1999**, *12*, 130–133.
- Issartel, J.P. Rebuilding sources of linear tracers after atmospheric concentration measurements. *Atmos. Chem. Phys.* **2003**, *3*, 2111–2125. [[CrossRef](#)]
- Mamonov, A.V.; Tsai, Y.H.R. Point source identification in nonlinear advection-diffusion-reaction systems. *Inverse Probl.* **2013**, *29*, 035009. [[CrossRef](#)]
- Turbelin, G.; Singh, S.K.; Issartel, J.P. Reconstructing source terms from atmospheric concentration measurements: Optimality analysis of an inversion technique. *J. Adv. Model. Earth Syst.* **2014**, *6*, 1244–1255. [[CrossRef](#)]

22. Kumar, P.; Feiz, A.A.; Singh, S.K.; Ngae, P.; Turbelin, G. Reconstruction of an atmospheric tracer source in an urban-like environment. *J. Geophys. Res. Atmos.* **2015**, *120*, 12589–12604. [[CrossRef](#)]
23. Bieringer, P.E.; Young, G.S.; Rodriguez, L.M.; Annunzio, A.J.; Vandenberghe, F.; Haupt, S.E. Paradigms and commonalities in atmospheric source term estimation methods. *Atmos. Environ.* **2017**, *156*, 102–112. [[CrossRef](#)]
24. Bocquet, M.; Elbern, H.; Eskes, H.; Hirtl, M.; Žabkar, R.; Carmichael, G.R.; Flemming, J.; Inness, A.; Pagowski, M.; Camaño, J.L.P.; et al. Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models. *Atmos. Chem. Phys. Discuss.* **2014**, *14*, 32233–32323. [[CrossRef](#)]
25. Carrassi, A.; Bocquet, M.; Bertino, L.; Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev. Clim. Chang.* **2018**, *9*, e535. [[CrossRef](#)]
26. Elbern, H.; Friese, E.; Nieradzik, L.; Schwinger, J. Data assimilation in atmospheric chemistry and air quality. In *Advanced Data Assimilation for Geosciences*; Oxford University Press: Oxford, UK, 2014; pp. 507–534. [[CrossRef](#)]
27. Silver, J.D.; Christensen, J.H.; Kahnert, M.; Robertson, L.; Rayner, P.J.; Brandt, J. Multi-species chemical data assimilation with the Danish Eulerian hemispheric model: system description and verification. *J. Atmos. Chem.* **2015**, *73*, 261–302. [[CrossRef](#)]
28. Nguyen, C.; Soulhac, L.; Salizzoni, P. Source Apportionment and Data Assimilation in Urban Air Quality Modelling for NO<sub>2</sub>: The Lyon Case Study. *Atmosphere* **2018**, *9*, 8. [[CrossRef](#)]
29. Xing, J.; Li, S.; Ding, D.; Kelly, J.T.; Wang, S.; Jang, C.; Zhu, Y.; Hao, J. Data Assimilation of Ambient Concentrations of Multiple Air Pollutants Using an Emission-Concentration Response Modeling Framework. *Atmosphere* **2020**, *11*, 1289. [[CrossRef](#)] [[PubMed](#)]
30. Mijling, B. High-resolution mapping of urban air quality with heterogeneous observations: A new methodology and its application to Amsterdam. *Atmos. Meas. Tech.* **2020**, *13*, 4601–4617. [[CrossRef](#)]
31. Nguyen, C.V.; Soulhac, L. Data assimilation methods for urban air quality at the local scale. *Atmos. Environ.* **2021**, *253*, 118366. [[CrossRef](#)]
32. Elbern, H.; Strunk, A.; Schmidt, H.; Talagrand, O. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys. Discuss.* **2007**, *7*, 1725–1783. [[CrossRef](#)]
33. Huang, W.S.; Griffith, S.M.; Lin, Y.C.; Chen, Y.C.; Lee, C.T.; Chou, C.C.K.; Chuang, M.T.; Wang, S.H.; Lin, N.H. Satellite-based Emission Inventory Adjustments Improve Simulations of Long-range Transport Events. *Aerosol Air Qual. Res.* **2021**, *21*, 210121. [[CrossRef](#)]
34. Markakis, K.; Valari, M.; Perrussel, O.; Sanchez, O.; Honore, C. Climate-forced air-quality modeling at the urban scale: sensitivity to model resolution, emissions and meteorology. *Atmos. Chem. Phys.* **2015**, *15*, 7703–7723. [[CrossRef](#)]
35. Holnicki, P.; Nahorski, Z. Emission Data Uncertainty in Urban Air Quality Modeling—Case Study. *Environ. Model. Assess.* **2015**, *20*, 583–597. [[CrossRef](#)]
36. Munir, S.; Mayfield, M.; Coca, D. Understanding Spatial Variability of NO<sub>2</sub> in Urban Areas Using Spatial Modelling and Data Fusion Approaches. *Atmosphere* **2021**, *12*, 179. [[CrossRef](#)]
37. Ponomarev, N.; Yushkov, V.; Elansky, N. Air Pollution in Moscow Megacity: Data Fusion of the Chemical Transport Model and Observational Network. *Atmosphere* **2021**, *12*, 374. [[CrossRef](#)]
38. Carnevale, C.; Angelis, E.D.; Finzi, G.; Turrini, E.; Volta, M. Application of Data Fusion Techniques to Improve Air Quality Forecast: A Case Study in the Northern Italy. *Atmosphere* **2020**, *11*, 244. [[CrossRef](#)]
39. Penenko, V.V.; Penenko, A.V.; Tsvetova, E.A.; Gochakov, A.V. Methods for Studying the Sensitivity of Air Quality Models and Inverse Problems of Geophysical Hydrothermodynamics. *J. Appl. Mech. Tech. Phys.* **2019**, *60*, 392–399. [[CrossRef](#)]
40. Penenko, A. A Newton-Kantorovich Method in Inverse Source Problems for Production-Destruction Models with Time Series-Type Measurement Data. *Numer. Anal. Appl.* **2019**, *12*, 51–69. [[CrossRef](#)]
41. Penenko, A. Convergence analysis of the adjoint ensemble method in inverse source problems for advection-diffusion-reaction models with image-type measurements. *Inverse Probl. Imaging* **2020**, *14*, 757–782. [[CrossRef](#)]
42. Penenko, A.; Zubairova, U.; Mukatova, Z.; Nikolaev, S. Numerical algorithm for morphogen synthesis region identification with indirect image-type measurement data. *J. Bioinform. Comput. Biol.* **2019**, *17*, 1940002–1–1940002–18. [[CrossRef](#)]
43. Penenko, A.; Gochakov, A.; Penenko, V. Algorithms based on sensitivity operators for analyzing and solving inverse modeling problems of transport and transformation of atmospheric pollutants. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *611*, 012032. [[CrossRef](#)]
44. Khodzher, T.V.; Zhamsueva, G.S.; Zayakhanov, A.S.; Dementeva, A.L.; Tsydygov, V.V.; Balin, Y.S.; Penner, I.E.; Kokhanenko, G.P.; Nasonov, S.V.; Klemasheva, M.G.; Golobokova, L.P.; Potemkin, V.L. Ship-Based Studies of Aerosol-Gas Admixtures over Lake Baikal Basin in Summer 2018. *Atmos. Ocean. Opt.* **2019**, *32*, 434–441. [[CrossRef](#)]
45. Antokhin, P.N.; Arshinova, V.G.; Arshinov, M.Y.; Belan, B.D.; Belan, S.B.; Davydov, D.K.; Ivlev, G.A.; Fofonov, A.V.; Kozlov, A.V.; Paris, J.D.; et al. Distribution of Trace Gases and Aerosols in the Troposphere Over Siberia During Wildfires of Summer 2012. *J. Geophys. Res. Atmos.* **2018**, *123*, 2285–2297. [[CrossRef](#)]
46. Gu, Q.; Michanowicz, D.R.; Jia, C. Developing a Modular Unmanned Aerial Vehicle (UAV) Platform for Air Pollution Profiling. *Sensors* **2018**, *18*, 4363. [[CrossRef](#)] [[PubMed](#)]
47. Arshinov, M.Y.; Belan, B.D.; Davydov, D.K.; Kozlov, A.V.; Fofonov, A.V.; Arshinova, V.G. Study of the Spatial Distributions of CO<sub>2</sub> and CH<sub>4</sub> in the Surface Air Layer over Western Siberia Using a Mobile Platform. *Atmos. Ocean. Opt.* **2020**, *33*, 661–670. [[CrossRef](#)]
48. Roshydromet. Unified Information System for Monitoring Atmospheric Air Pollution. Available online: <http://www.feerc.ru/uisem/portal/ad/irkutsk> (accessed on 1 November 2021). (In Russian)

49. Marchuk, G.I. Formulation of some converse problems. *Sov. Math. Dokl.* **1964**, *5*, 675–678.
50. Penenko, V. *Methods for Numerical Simulation of Atmospheric Processes*; Hydrometeoizdat: Leningrad, Russia, 1981. (In Russian)
51. Murio, D.A. *The Mollification Method and the Numerical Solution of Ill-Posed Problems*; John Wiley & Sons, Inc.: New York, NY, USA, 1993. [[CrossRef](#)]
52. Dimet, F.X.L.; Souopgui, I.; Titaud, O.; Shutyaev, V.; Hussaini, M.Y. Toward the assimilation of images. *Nonlinear Process. Geophys.* **2015**, *22*, 15–32. [[CrossRef](#)]
53. Penenko, V.V.; Obratsov, N.N. A variational initialization method for the fields of the meteorological elements. *Engl. Transl. Sov. Meteorol. Hydrol.* **1976**, *11*, 3–16.
54. Marchuk, G.I.; Penenko, V.V. Application of optimization methods to the problem of mathematical simulation of atmospheric processes and environment. In *Modelling and Optimization of Complex System*; Springer: Berlin/Heidelberg, Germany, 1978; pp. 240–252. [[CrossRef](#)]
55. Dimet, F.X.L.; Talagrand, O. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **1986**, *38A*, 97–110. [[CrossRef](#)]
56. Issartel, J.P. Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation. *Atmos. Chem. Phys.* **2005**, *5*, 249–273. [[CrossRef](#)]
57. Penenko, A.V.; Gochakov, A.; Antokhin, P. Numerical study of emission sources identification algorithm with joint use of in situ and remote sensing measurement data. In Proceedings of the 26th International Symposium on Atmospheric and Ocean Optics, Atmospheric Physics, Moscow, Russia, 29 June–3 July 2020; Matvienko, G.G.; Romanovskii, O.A., Eds.; SPIE: Bellingham, WA, USA, 2020. [[CrossRef](#)]
58. UNESCO. Lake Baikal. Available online: <https://whc.unesco.org/en/list/754/> (accessed on 1 December 2021).
59. Plyusnin, V.; Vladimirov, I.; Sorokovoi, A. Baikal region in the UNESCO “Man and Biocphere” Programme. *Probl. Geogr.* **2021**, *152*, 202–221. [[CrossRef](#)]
60. Golobokova, L.; Khodzher, T.; Khuriganova, O.; Marinayte, I.; Onishchuk, N.; Rusanova, P.; Potemkin, V. Variability of Chemical Properties of the Atmospheric Aerosol above Lake Baikal during Large Wildfires in Siberia. *Atmosphere* **2020**, *11*, 1230. [[CrossRef](#)]
61. Efimova, N.V.; Rukavishnikov, V.S. Assessment of Smoke Pollution Caused by Wildfires in the Baikal Region (Russia). *Atmosphere* **2021**, *12*, 1542. [[CrossRef](#)]
62. Popovicheva, O.; Molozhnikova, E.; Nasonov, S.; Potemkin, V.; Penner, I.; Klemasheva, M.; Marinaite, I.; Golobokova, L.; Vratolis, S.; Eleftheriadis, K.; et al. Industrial and wildfire aerosol pollution over world heritage Lake Baikal. *J. Environ. Sci.* **2021**, *107*, 49–64. [[CrossRef](#)]
63. Gorshkov, A.G.; Izosimova, O.N.; Kustova, O.V.; Marinaite, I.I.; Galachyants, Y.P.; Sinyukovich, V.N.; Khodzher, T.V. Wildfires as a Source of PAHs in Surface Waters of Background Areas (Lake Baikal, Russia). *Water* **2021**, *13*, 2636. [[CrossRef](#)]
64. Grigorieva, M.A.; Ippolitova, N.A. Big business in socio-economic development of cities in the Baikal region. *Geogr. Nat. Resour.* **2011**, *32*, 166–171. [[CrossRef](#)]
65. Akhtimankina, A.; Arguchintseva, A. Zagryaznenie atmosfernogo vozduha promyshlennymi predpriyatiyami g. Irkutsk. *IZVESTIYA Irkutsk. Gos. Univ.* **2013**, *6*, 3–19. (In Russian)
66. Brown, K.P.; Gerber, A.; Bedulina, D.; Timofeyev, M.A. Human impact and ecosystemic health at Lake Baikal. *WIREs Water* **2021**, *8*, e1528. [[CrossRef](#)]
67. Khuriganova, O.I.; Obolkin, V.A.; Golobokova, L.P.; Bukin, Y.S.; Khodzher, T.V. Passive Sampling as a Low-Cost Method for Monitoring Air Pollutants in the Baikal Region (Eastern Siberia). *Atmosphere* **2019**, *10*, 470. [[CrossRef](#)]
68. Zayakhanov, A.S.; Zhamsueva, G.S.; Tcydypov, V.V.; Balzhanov, T.S.; Dementeva, A.L.; Khodzher, T.V. Investigation of Transport and Transformation of Tropospheric Ozone in Terrestrial Ecosystems of the Coastal Zone of Lake Baikal. *Atmosphere* **2019**, *10*, 739. [[CrossRef](#)]
69. Mashyanov, N.; Obolkin, V.; Pogarev, S.; Ryzhov, V.; Sholupov, S.; Potemkin, V.; Molozhnikova, E.; Khodzher, T. Air Mercury Monitoring at the Baikal Area. *Atmosphere* **2021**, *12*, 807. [[CrossRef](#)]
70. Golobokova, L.; Netsvetaeva, O.; Khodzher, T.; Obolkin, V.; Khuriganova, O. Atmospheric Deposition on the Southwest Coast of the Southern Basin of Lake Baikal. *Atmosphere* **2021**, *12*, 1357. [[CrossRef](#)]
71. Obolkin, V.; Molozhnikova, E.; Shikhovtsev, M.; Netsvetaeva, O.; Khodzher, T. Sulfur and Nitrogen Oxides in the Atmosphere of Lake Baikal: Sources, Automatic Monitoring, and Environmental Risks. *Atmosphere* **2021**, *12*, 1348. [[CrossRef](#)]
72. Engl, H.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*; Kluwer: Dordrecht, The Netherlands, 1996.
73. Voronina, T.A.; Tcheverda, V.A.; Voronin, V.V. Some properties of the inverse operator for a tsunami source recovery. *Sib. Elektron. Mat. Izv.* **2014**, *11*, 532–547.
74. Judd, L.M.; Al-Saadi, J.A.; Valin, L.C.; Pierce, R.B.; Yang, K.; Janz, S.J.; Kowalewski, M.G.; Szykman, J.J.; Tiefengraber, M.; Mueller, M. The Dawn of Geostationary Air Quality Monitoring: Case Studies From Seoul and Los Angeles. *Front. Environ. Sci.* **2018**, *6*, 1–17. [[CrossRef](#)]
75. Kim, J.; Jeong, U.; Ahn, M.H.; Kim, J.H.; Park, R.J.; Lee, H.; Song, C.H.; Choi, Y.S.; Lee, K.H.; Yoo, J.M.; et al. New Era of Air Quality Monitoring from Space: Geostationary Environment Monitoring Spectrometer (GEMS). *Bull. Am. Meteorol. Soc.* **2020**, *101*, E1–E22. [[CrossRef](#)]

76. Mettig, N.; Weber, M.; Rozanov, A.; Arosio, C.; Burrows, J.P.; Veefkind, P.; Thompson, A.M.; Querel, R.; Leblanc, T.; Godin-Beekmann, S.; et al. Ozone profile retrieval from nadir TROPOMI measurements in the UV range. *Atmos. Meas. Tech.* **2021**, *14*, 6057–6082. [[CrossRef](#)]
77. Liu, S.; Valks, P.; Pinardi, G.; Xu, J.; Chan, K.L.; Argyrouli, A.; Lutz, R.; Beirle, S.; Khorsandi, E.; Baier, F.; et al. An improved TROPOMI tropospheric  $\text{NO}_2$  research product over Europe. *Atmos. Meas. Tech.* **2021**, *14*, 7297–7327. [[CrossRef](#)]
78. Stebel, K.; Stachlewska, I.S.; Nemuc, A.; Horálek, J.; Schneider, P.; Ajtai, N.; Diamandi, A.; Benešová, N.; Boldeanu, M.; Botezan, C.; et al. SAMIRA-SATellite Based Monitoring Initiative for Regional Air Quality. *Remote Sens.* **2021**, *13*, 2219. [[CrossRef](#)]
79. Wolfram Research. Wolfram Alpha. Available online: <https://www.wolframalpha.com/> (accessed on 12 December 2020).
80. Hundsdorfer, W.; Verwer, J.G. *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*; Springer Series in Computational Mathematics; Springer: Berlin/Heidelberg, Germany, 2013.
81. Hauglustaine, D.A.; Brasseur, G.P.; Walters, S.; Rasch, P.J.; Muller, J.F.; Emmons, L.K.; Carroll, M.A. MOZART, a global chemical transport model for ozone and related chemical tracers: 2. Model results and evaluation. *J. Geophys. Res. Atmos.* **1998**, *103*, 28291–28335. [[CrossRef](#)]
82. Baldauf, M.; Seifert, A.; Forstner, J.; Majewski, D.; Raschendorfer, M.; Reinhardt, T. Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Mon. Weather. Rev.* **2011**, *139*, 3887–3905. [[CrossRef](#)]
83. Griewank, A.; Walther, A. *Evaluating Derivatives*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2008. [[CrossRef](#)]
84. Vlasenko, A.; Kohl, A.; Stammer, D. The efficiency of geophysical adjoint codes generated by automatic differentiation tools. *Comput. Phys. Commun.* **2016**, *199*, 22–28. [[CrossRef](#)]
85. Naumann, U. Adjoint Code Design Patterns. *ACM Trans. Math. Softw.* **2019**, *45*, 1–32. [[CrossRef](#)]
86. Penenko, A.; Gochakov, A. Parallel speedup analysis of an adjoint ensemble-based source identification algorithm. *J. Phys. Conf. Ser.* **2021**, *1715*, 012072–1–012072–6. [[CrossRef](#)]
87. Koh, J.; Lee, J.; Yoon, S. Single-image deblurring with neural networks: A comparative survey. *Comput. Vis. Image Underst.* **2021**, *203*, 103134. [[CrossRef](#)]
88. Zhang, Q.; Hu, Z.; Jiang, C.; Zheng, H.; Ge, Y.; Liang, D. Artifact removal using a hybrid-domain convolutional neural network for limited-angle computed tomography imaging. *Phys. Med. Biol.* **2020**, *65*, 155010. [[CrossRef](#)]
89. Xie, S.; Zheng, X.; Chen, Y.; Xie, L.; Liu, J.; Zhang, Y.; Yan, J.; Zhu, H.; Hu, Y. Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction. *Sci. Rep.* **2018**, *8*. [[CrossRef](#)]
90. Penenko, A.; Nikolaev, S.; Golushko, S.; Romashenko, A.; Kirilova, I. Numerical Algorithms for Diffusion Coefficient Identification in Problems of Tissue Engineering. *Math. Biol. Bioinform.* **2016**, *11*, 426–444. (In Russian), [[CrossRef](#)]
91. Cheverda, V.A.; Kostin, V.I. R-pseudoinverses for compact operators in Hilbert spaces: existence and stability. *J. Inverse Ill-Posed Probl.* **1995**, *3*, 131–148. [[CrossRef](#)]
92. Penenko, A.V.; Salimova, A.B. Source Identification for the Smoluchowski Equation Using an Ensemble of Adjoint Equation Solutions. *Numer. Anal. Appl.* **2020**, *13*, 152–164. [[CrossRef](#)]